Multi-Server Weakly-Private Information Retrieval

Hsuan-Yin Lin, *Senior Member, IEEE*, Siddhartha Kumar,
Eirik Rosnes, *Senior Member, IEEE*, Alexandre Graell i Amat, *Senior Member, IEEE*,
and Eitan Yaakobi, *Senior Member, IEEE*

Abstract— Private information retrieval (PIR) protocols ensure that a user can download a file from a database without revealing any information on the identity of the requested file to the servers storing the database. While existing protocols strictly impose that no information is leaked on the file's identity, this work initiates the study of the tradeoffs that can be achieved by relaxing the perfect privacy requirement. We refer to such protocols as weakly-private information retrieval (WPIR) protocols. In particular, for the case of multiple noncolluding replicated servers, we study how the download rate, the upload cost, and the access complexity can be improved when relaxing the perfect privacy constraint. To quantify the information leakage on the requested file's identity we consider mutual information (MI), worst-case information leakage, and maximal leakage (MaxL). We present two WPIR schemes, denoted by Scheme A and Scheme B, based on two recent PIR protocols and show that the download rate of the former can be optimized by solving a convex optimization problem. We also show that Scheme A achieves an improved download rate compared to the recently proposed scheme by Samy *et al.* under the so-called ϵ-privacy metric. Additionally, a family of schemes based on partitioning is presented. Moreover, we provide an information-theoretic converse bound for the maximum possible download rate for the MI and MaxL privacy metrics under a practical restriction on the alphabet size of queries and answers. For two servers and two files, the bound is tight under the MaxL metric, which settles the WPIR capacity in this particular case. Finally, we compare the performance of the proposed schemes and their gap to the converse bound.

Index Terms— Capacity, information leakage, information-theoretic privacy, multiple servers, private information retrieval.

I. INTRODUCTION

PRIVATE information retrieval (PIR) was introduced in the computer science literature by Chor *et al.* in [1] and [2].

A PIR scheme allows a user to retrieve an arbitrary file from a database that is stored on either a single or multiple servers without revealing any information about the identity of the requested file. The efficiency of a PIR scheme is measured in terms of the total communication load, consisting of both the upload and download cost for the retrieval of a single file. It was already shown in the original work of Chor *et al.* [2] that in the case that the database is stored on a single server, all files need to be downloaded in order to achieve perfect privacy, i.e., no information leakage on the identity of the requested file. It has been extensively studied how to reduce the communication cost using several copies of the database, see, e.g., [3]–[6].

From an information-theoretic perspective and for many practical scenarios, the file size is typically much larger than the size of the queries to all servers. Therefore, rather than accounting for both the upload and the download cost, as usually done in the computer science literature, in the information theory literature efficiency is typically measured in terms of the download cost, or equivalently, in terms of the download rate. The download rate—or the PIR rate—is defined as the ratio between the requested file size and the average number of downloaded symbols for the retrieval of a single file. The maximum possible PIR rate of all possible schemes is called the PIR capacity. The PIR capacity for the classical PIR model of replicated servers was characterized by Sun and Jafar [7].

To achieve a lower storage overhead, PIR protocols have also been considered jointly with coded distributed storage systems (DSSs), where the data is encoded by a linear code and then stored on several servers in a distributed manner [8]–[10]. The case of maximum distance separable (MDS) coded servers was considered in [11]–[16], while the case of arbitrary linear coded servers was studied in [17]–[20]. The concept of PIR has also been extended to several other relevant scenarios, which include colluding servers [11], [17], [19], [21]–[24], robust PIR with Byzantine or unresponsive servers [21], [25], [26], multi-round PIR [27], multi-file PIR [28], optimal download cost of PIR for an arbitrary file size [29], optimal upload cost of PIR, i.e., the minimum required amount of query information [30], access complexity of PIR, i.e., the number of symbols accessed across all servers for the retrieval of a single file [31], tradeoff between the storage and download cost of PIR [32], cache-aided PIR [33], [34], PIR with side information [35]–[39], PIR on graph-based replication systems [40], [41], PIR with secure storage [42]–[44], functional PIR codes [45], and private proximity retrieval codes [46].

All of the aforementioned extensions of PIR impose perfect privacy, i.e., no information leakage. However, this assumption

is quite restrictive and may be relaxed for several practical applications, as leaking part of the information of the identity of the requested file is legitimate as long as there is still enough ambiguity on the file's identity to meet the privacy requirement specified by the user. For example, the user may be willing to share with the servers that the file is a movie (and not a book or other forms of files), or only the movie's genre, whereas keeping private the identity of the movie. Relaxing the perfect privacy requirement of PIR has been considered briefly in the computer science literature previously. As early as in 2002, Asonov *et al.* [47] introduced the concept of repudiation as a relaxation of PIR. Their main motivation was to reduce the preprocessing complexity of queries, while keeping optimal communication (upload and download) cost and response time. However, the condition of repudiation can be achieved even if the server can determine the identity of the requested file almost surely. Hence, it does not provide a good level of information-theoretic privacy. More than a decade later, Toledo *et al.* [48] adopted a privacy metric based on *differential privacy* [49], [50] and traded off privacy for reduced communication cost. In [48], several schemes that hide the query identity were proposed and studied. However, different mechanisms and security assumptions from those of information-theoretic PIR were considered. Interestingly, the authors claim that the proposed approaches can be applied to information-theoretic PIR, but they did not study any fundamental information-theoretic tradeoffs between information leakage and different costs under the considered privacy metric.

This paper takes a first step away from the perfect privacy requirement of the information-theoretic PIR framework. Our goal is to study the tradeoffs between different parameters of PIR, such as download rate, upload cost, and access complexity, while relaxing the perfect privacy requirement on the identity of the desired file. We refer to such a scenario as *weakly-private* information retrieval (WPIR). How to properly measure information leakage has been studied extensively in the computer science literature, see, e.g., [51] and references therein. Mutual information (MI) [52]–[54], that captures the *average* information leakage between the private data and the adversary's observations, maximal leakage (MaxL) [55], [56], and worst-case information leakage (WIL) [57], are among the most popular information-theoretic privacy leakage metrics, along with (local) differential privacy [49], [50], [58], [59]. To the best of our knowledge, using MI as a privacy metric originates from the domain of genome privacy and was first considered in [52]. Although, the MI privacy metric has a less clear operational meaning than MaxL, the presented results for the MI privacy metric provide fundamental insight into the tradeoff between download cost and privacy leakage, which is also valid and complements the presented results for the other considered privacy metrics. In this work, we consider the case of replicated noncolluding servers, mainly focusing on the MI and MaxL privacy metrics. We propose a WPIR scheme by building upon a PIR protocol recently introduced in [30] and study its tradeoffs between download rate, upload cost, and access complexity. In particular, we show that by relaxing the perfect privacy requirement, the download rate can be improved beyond PIR capacity.

The main contributions can be summarized as follows:

- We introduce the concept of an (M, n) information retrieval (IR) scheme for a DSS with n servers storing M files using a global random strategy vector and a corresponding scheme, referred to as Scheme A, by building upon a PIR protocol introduced in [30]. By selecting each entry of the global random strategy according to a Bernoulli distribution, we provide for the special case of $n = 2$ servers closed-form expressions for the achievable download rate, upload cost, access complexity, and privacy leakage (see Theorem 3).

- In addition, we adopt the privacy metric introduced in the related works [60], [61], the so-called ϵ-privacy, and compare our proposed Scheme A to their leaky PIR scheme. By using a global random strategy for which each entry is independent and identically distributed (i.i.d.) according to a Bernoulli distribution, we show that Scheme A performs better in terms of download rate for the case of $n = 2$ servers.

- By using a time-sharing argument (see Theorem 1 and the discussion in Section VII), the download rate of Scheme A can be improved. For both the MI and MaxL privacy metrics we show that optimizing the download rate for Scheme A with time-sharing over the global random strategy can be framed as a convex optimization problem (see Section VII).

- We provide an information-theoretic converse result for the maximum possible download rate for an (M, n) IR scheme for both the MI and MaxL privacy metrics in Theorems 8 and 10, respectively, under a practical restriction on the alphabet size of queries and answers. The converse is derived using a known result between the entropy difference and the total variation (TV) distance of two probability distributions (see Lemma 5). For the special case of $(M, n) = (2, 2)$ the WPIR capacity is provided in Theorems 9 (assuming that only one of the two servers can leak) and 11 for the MI and MaxL privacy metrics, respectively. Moreover, we show that Scheme A with time-sharing and with each entry of the global random strategy selected according to a Bernoulli distribution achieves the WPIR capacity for both privacy metrics under the above restrictions in this special case.

- Extensive numerical results showing the tradeoff between download rate, upload cost, access complexity, and privacy leakage are presented in Section X for Scheme A (with and without time-sharing). As a comparison, we also compare with an alternative proposed constant-rate IR scheme, referred to as Scheme B and based on the PIR scheme in [17, Lem. 4].

A. Related Work

Independently, the download rate-leakage tradeoff has been studied by Samy *et al.* [60] under the name of leaky PIR using a privacy metric related to differential privacy. The leaky PIR

framework was recently also extended to symmetric PIR [61] and to the consideration of latent attributes in the single server case [62]. Symmetric PIR is a variant of PIR where in addition the user cannot learn anything about the remaining files in the database when the user retrieves its desired file [63]–[65]. Moreover, Zhou *et al.* [66] have recently studied the same problem under the MaxL privacy metric. Their scheme builds upon the same PIR protocol as our proposed Scheme A. Moreover, by allowing for a permutation of the query strategy across the servers in addition to an arbitrary global random strategy, improved performance can be achieved. It can be shown that their scheme is equivalent to our Scheme A with time-sharing.

Our companion paper [67] studies the corresponding problem for the single server setting under both the MI and MaxL privacy metrics. In particular, by relating the WPIR problem to rate-distortion theory, the capacity of single-server WPIR is fully characterized. Lastly, the related work in [68] is also worth mentioning. In contrast to WPIR, where information leakage on the identity of the desired file to the servers is considered, the information leakage of the nondesired files to the user for classical PIR was studied in [68].

B. Organization of Paper

The remainder of this paper is organized as follows. Section II presents the notation, basic definitions, preliminaries, and the problem formulation. In Section III, we present a partition scheme which first divides the database into equally-sized partitions and then uses a given IR scheme to retrieve a file from the corresponding partition. Section IV presents an IR scheme built upon the PIR protocol introduced in [30], referred to as Scheme A, while Section V presents a constant-rate IR scheme based on the PIR scheme in [17, Lem. 4], referred to as Scheme B. For both schemes and for two servers we provide closed-form expressions for the download rate, upload cost, access complexity, and information leakage under a Bernoulli global random strategy. In Section IV, using Scheme A to retrieve files from a partition of the partition scheme is also analyzed. Based on Scheme A, a WPIR scheme achieving a better download rate than the leaky PIR scheme under the ϵ-privacy metric [60], [61] is proposed in Section VI. The minimization of the information leakage for Scheme A with time-sharing is considered in Section VII. In particular, we show that the minimization problem is a convex optimization problem for both the MI and MaxL privacy metrics. Then, in Sections VIII and IX we present converse results on the minimum download cost for both privacy metrics under a practical restriction on the alphabet size of queries and answers. Numerical results comparing Schemes A (with and without time-sharing) and B in terms of download rate, upload cost, access complexity, and information leakage are presented in Section X. Finally, some conclusions are drawn in Section XI.

II. Preliminaries

A. Notations

We denote by $\mathbb{N}$ the set of all positive integers, $[a] \triangleq \{1, 2, \ldots, a\}$, and $[a : b] \triangleq \{a, a+1, \ldots, b\}$ for $a, b \in \{0\} \cup \mathbb{N}$, $a \leq b$. Vectors are denoted by bold letters, random variables (RVs) (either scalar or vector) by uppercase letters, and sets by calligraphic uppercase letters, e.g., $\boldsymbol{x}$, X, and $\mathcal{X}$, respectively. Moreover, $\mathcal{X}^{\mathrm{c}}$ denotes the complement of a set $\mathcal{X}$ in a universe set. The all-zero matrix of dimensions $a \times b$ is represented by $\mathbf{0}_{a \times b}$, or simply by $\mathbf{0}$ when the dimensions are not important. For a given index set $\mathcal{S}$, we write $X^{\mathcal{S}}$ and $Y_{\mathcal{S}}$ to represent $\{X^{(m)} : m \in \mathcal{S}\}$ and $\{Y_l : l \in \mathcal{S}\}$, respectively. $X \perp\!\!\!\perp Y$ means that the two RVs X and Y are independent. $\mathbb{E}_X [\cdot]$ denotes the expectation over the RV X. $X \sim \mathrm{Bernoulli}(p)$ denotes a Bernoulli-distributed RV with $\Pr[X = 1] = p = 1 - \Pr[X = 0]$ and $X \sim \mathcal{U}(\mathcal{S})$ a uniformly-distributed RV over the set $\mathcal{S}$. $(\cdot)^{\mathsf{T}}$ denotes the transpose of its argument. The Hamming weight of a binary vector $\boldsymbol{x}$ is denoted by $w_{\mathsf{H}}(\boldsymbol{x})$, while its support is denoted by $\chi(\boldsymbol{x})$. $\sigma(\cdot)$ denotes a left cyclic permutation, while a left cyclic shifts are obtained through function composition and denoted by $\sigma^a(\cdot)$. The inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ is denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$. $\mathsf{H}(X)$, $\mathsf{H}(P_X)$, or $\mathsf{H}(p_1, \ldots, p_{|\mathcal{X}|})$ represents the entropy of X, where $P_X(\cdot) = (p_1, \ldots, p_{|\mathcal{X}|})$ denotes the distribution of the RV X, while $\mathsf{I}(X; Y)$ is the MI between X and Y (in bits). $\mathsf{H}_{\mathsf{b}}(p) \triangleq -p \log_2 p - (1-p) \log_2 (1-p)$ is the binary entropy function. With some abuse of notation, when the marginal distribution of either X or Y is assumed fixed and known, the MI between X and Y is sometimes simply written as $\mathsf{I}(X; Y) \equiv \mathsf{I}(P_{X|Y}) \equiv \mathsf{I}(P_{Y|X})$.

B. System Model

We consider a DSS with n noncolluding replicated servers, each storing M independent files $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(\mathsf{M})}$, where each file $\boldsymbol{X}^{(m)} = (X_1^{(m)}, \ldots, X_\beta^{(m)})^{\mathsf{T}}$, $m \in [\mathsf{M}]$, has length β, and can be seen as a $\beta \times 1$ vector over an alphabet $\mathcal{X}$. Assume that each element of $\boldsymbol{X}^{(m)}$ is chosen independently and uniformly at random from $\mathcal{X}$. Thus, we have $\mathsf{H}(\boldsymbol{X}^{(m)}) = \beta \log_2 |\mathcal{X}|$, $\forall m \in [\mathsf{M}]$.

In information retrieval (IR), a user wishes to efficiently retrieve one of the M files stored in the replicated DSS. Similar to the detailed mathematical description in [30], we assume that the requested file index M is a RV and $M \sim \mathcal{U}([\mathsf{M}])$.[1] We give the following definition of an IR scheme.

Definition 1: An (M, n) IR scheme $\mathscr{C}$ for a DSS with n servers storing M files consists of:
- A global random strategy $\boldsymbol{S}$, whose alphabet is $\mathcal{S}$.
- n query-encoding functions ϕ_l, $l \in [n]$, that generate n queries $\boldsymbol{Q}_l = \phi_l(M, \boldsymbol{S})$ with alphabet $\mathcal{Q}_l$, where query $\boldsymbol{Q}_l$ is sent to server l.
- n answer-length functions $\ell_l(\boldsymbol{Q}_l)$, with range $\{0\} \cup \mathbb{N}$, that define the length of the answers. $\ell_l(\boldsymbol{Q}_l)$ is a function of the query $\boldsymbol{Q}_l$, which is independent of the particular realization of the files.
- n answer functions

$$\varphi_l \colon \mathcal{Q}_l \times \mathcal{X}^{\beta \mathsf{M}} \to \mathcal{A}^{\ell_l}, \quad l \in [n],$$

that return the answers $\boldsymbol{A}_l = \varphi_l(\boldsymbol{Q}_l, \boldsymbol{X}^{[\mathsf{M}]})$, where $\mathcal{A}$ is the download symbol alphabet.

---

[1] Here, we assume for simplicity that the requested file index M is uniformly distributed. However, this assumption can be lifted, which is referred to as semantic PIR in the literature [69].

- n access-number functions $\delta_l(\boldsymbol{Q}_l)$, with range $\{0\} \cup \mathbb{N}$, that define the number of symbols accessed by $\boldsymbol{Q}_l$.

This scheme should satisfy the condition of *perfect retrievability*,

$$\mathsf{H}\big(\boldsymbol{X}^{(M)} \mid \boldsymbol{A}_{[n]}, \boldsymbol{Q}_{[n]}, M\big) = 0. \qquad (1)$$

$\diamondsuit$

Since a user should be able to generate the queries without any prior knowledge of the realizations of the files, it is reasonable to assume that the queries and the files are independent, i.e.,

$$\mathsf{I}\big(\boldsymbol{X}^{[\mathsf{M}]}; \boldsymbol{Q}_{[n]}\big) = 0. \qquad (2)$$

This particular assumption is used in the converse proofs of Sections VIII and IX (see Appendix G).

Note that a PIR scheme is an (M, n) IR scheme that satisfies perfect privacy for all servers, i.e., for every $m, m' \in [\mathsf{M}]$ with $m \neq m'$, the condition

$$\Pr[\boldsymbol{Q}_l = \boldsymbol{q}_l | M = m] = \Pr[\boldsymbol{Q}_l = \boldsymbol{q}_l | M = m'] \qquad (3)$$

holds for all $\boldsymbol{q}_l \in \mathcal{Q}_l$, $l \in [n]$. The privacy constraint (3) is equivalent to the statement that $M \perp\!\!\!\perp \boldsymbol{Q}_l$. We denote by $\boldsymbol{Q}_l^{(m)}$ the query sent to server l if file $\boldsymbol{X}^{(m)}$ is requested, which is a RV with probability mass function (PMF) $P_{\boldsymbol{Q}_l^{(m)}}(\boldsymbol{q}_l) \triangleq \Pr[\boldsymbol{Q}_l = \boldsymbol{q}_l | M = m] = P_{\boldsymbol{Q}_l | M}(\boldsymbol{q}_l | m)$.

We refer to an (M, n) IR scheme that does not satisfy (3) as a *WPIR scheme*, as opposed to a PIR scheme that leaks no information.

C. Metrics of Information Leakage

In this paper, to measure the information leakage between M and $\boldsymbol{Q}_l$ for an IR scheme, we first consider MI, which measures the *average* amount of information about the requested file index M from the queries $\boldsymbol{Q}_l$. Moreover, we also consider WIL [57] and MaxL, which is considered a robust information leakage quantity [56]. For the MI privacy metric, we use the following theorem to motivate the definition of information leakage for an (M, n) IR scheme.

Theorem 1 (Time-Sharing Principle for the MI Metric): Consider an (M, n) IR scheme $\mathscr{C}$, where the leakage of the l-th server is defined as $\mathsf{I}(M; \boldsymbol{Q}_l)$, $l \in [n]$. Then, there exists an (M, n) IR scheme $\overline{\mathscr{C}}$ with leakage $\bar{\rho} \triangleq \frac{1}{n} \sum_{l \in [n]} \mathsf{I}(M; \boldsymbol{Q}_l)$ for every server. $\diamondsuit$

Proof: The theorem is proven by a time-sharing argument. Assume that the IR scheme $\mathscr{C}$ is given by the query-encoding functions ϕ_l, answer functions φ_l, $l \in [n]$, and a random strategy $\boldsymbol{S}$.

Next, define query-encoding functions $\bar{\phi}_l$, answer functions $\bar{\varphi}_l$, $l \in [n]$, and a random strategy $\boldsymbol{S}_T$ for an (M, n) IR scheme $\overline{\mathscr{C}}$ as follows. Given a requested file index M, the user chooses a $T \sim \mathcal{U}([n])$ and assigns the query $\overline{\boldsymbol{Q}}_l = \bar{\phi}_l(M, \boldsymbol{S}_T) \triangleq \phi_{\sigma^{T-1}(l)}(M, \boldsymbol{S}) = \boldsymbol{Q}_{\sigma^{T-1}(l)}(M, \boldsymbol{S})$ to the l-th server, $l \in [n]$.

The answer functions for $\overline{\mathscr{C}}$ are defined as $\bar{\varphi}_l(\overline{\boldsymbol{Q}}_l, \boldsymbol{X}^{[\mathsf{M}]}) \triangleq \varphi_{\sigma^{T-1}(l)}(\phi_{\sigma^{T-1}(l)}(M, \boldsymbol{S}), \boldsymbol{X}^{[\mathsf{M}]})$, $l \in [n]$, and hence perfect retrievability is achieved due to the perfect retrievability of the IR scheme $\mathscr{C}$.

The MI information leakage of the l-th server is

$$
\begin{aligned}
\mathsf{I}(M; \overline{\boldsymbol{Q}}_l) &= \mathsf{H}(\overline{\boldsymbol{Q}}_l) - \mathsf{H}(\overline{\boldsymbol{Q}}_l \mid M) \\
&= \mathsf{H}(\boldsymbol{Q}_{\sigma^{T-1}(l)}) - \mathsf{H}(\boldsymbol{Q}_{\sigma^{T-1}(l)} \mid M) \\
&\overset{(a)}{=} \sum_{t=1}^{n} \Pr[T = t]\, \mathsf{I}(\boldsymbol{Q}_{\sigma^{t-1}(l)}; M | T = t) \\
&= \frac{1}{n} \sum_{l'=1}^{n} \mathsf{I}(\boldsymbol{Q}_{l'}; M), \quad \forall\, l \in [n],
\end{aligned}
$$

where (a) follows from the definition of conditional mutual information. ∎

Theorem 1 indicates that we can always obtain an (M, n) IR scheme with equal MI leakage at each server by cyclically shifting the servers' queries of an existing (M, n) IR scheme $\mathscr{C}$ n times. Such a time-sharing scheme is denoted by $\overline{\mathscr{C}}$.

Hence, to characterize the overall leakage of a given (M, n) IR scheme $\mathscr{C}$ in terms of MI, we consider the information leakage metric

$$\rho^{(\mathsf{MI})}(\mathscr{C}) \triangleq \frac{1}{n} \sum_{l \in [n]} \mathsf{I}(M; \boldsymbol{Q}_l). \qquad (4)$$

The WIL of the l-th server is defined as $\mathsf{WIL}(M; \boldsymbol{Q}_l) \triangleq \mathsf{H}(M) - \min_{\boldsymbol{q}_l \in \mathcal{Q}_l} \mathsf{H}(M | \boldsymbol{Q}_l = \boldsymbol{q}_l)$. The overall WIL of a given (M, n) IR scheme $\mathscr{C}$ is then given as $\rho^{(\mathsf{WIL})}(\mathscr{C}) \triangleq \max_{l \in [n]} \mathsf{WIL}(M; \boldsymbol{Q}_l)$.

Further, given a joint distribution $P_{M, \boldsymbol{Q}}$, the MaxL from M to $\boldsymbol{Q}$ is defined as

$$\mathsf{MaxL}(M; \boldsymbol{Q}) \triangleq \log_2 \sum_{\boldsymbol{q} \in \mathcal{Q}} \max_{m \in [\mathsf{M}]} P_{\boldsymbol{Q}|M}(\boldsymbol{q}|m). \qquad (5)$$

Note that MaxL has a strong connection to the min-entropy (MinE) privacy metric, which is commonly-used in the computer science literature [55], [70]. MinE is a special case of the widely known Rényi entropy [71]. The MinE information leakage and the MaxL privacy metric can be shown to be equivalent when M is uniformly distributed [56], [70].

We will use (5) as the MaxL privacy metric for the designed query distribution $P_{\boldsymbol{Q}_l|M}$ at the l-th server of a WPIR scheme, which is denoted by

$$\rho^{(\mathsf{MaxL})}(M, \boldsymbol{Q}_l) \triangleq \mathsf{MaxL}(M; \boldsymbol{Q}_l).$$

The overall MaxL of a given (M, n) IR scheme $\mathscr{C}$ is then defined to be the worst-case MaxL over all servers:

$$\rho^{(\mathsf{MaxL})}(\mathscr{C}) \triangleq \max_{l \in [n]} \mathsf{MaxL}(M; \boldsymbol{Q}_l).$$

The following lemma summarizes some useful properties for both the MI and MaxL privacy metrics. $\diamondsuit$

Lemma 1 ([56, Lem. 1, Cor. 1]): For any joint distribution $P_{X,Y}$, we have the following.

1) (Data Processing Inequalities) If the RVs X, Y, and Z form a Markov chain, then

$$
\begin{aligned}
\mathsf{I}(X; Z) &\le \min\{\mathsf{I}(X; Y), \mathsf{I}(Y; Z)\}, \text{ and} \\
\mathsf{MaxL}(X; Z) &\le \min\{\mathsf{MaxL}(X; Y), \mathsf{MaxL}(Y; Z)\}.
\end{aligned}
$$

2) Consider a fixed distribution P_X. Then, both $\mathsf{I}(X; Y)$ and $2^{\mathsf{MaxL}(X; Y)}$ are convex functions in $P_{Y|X}$. $\diamondsuit$

There are other privacy metrics that can be used to relax the perfect privacy requirement of PIR. The authors of [60], [61] introduced the ϵ-*privacy* metric based on the notion of (local) differential privacy. Under the setup discussed in this paper, we define the ϵ-privacy leakage at the l-th server of an IR scheme $\mathscr{C}$ as

$$\rho^{(\epsilon\text{-P})}(M; \boldsymbol{Q}_l) \triangleq \ln \left(\max_{\boldsymbol{q}_l \in \mathcal{Q}_l} \max_{m, m' \in [\mathsf{M}]} \frac{P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)}{P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m')} \right).$$

Similar to the MaxL privacy metric, we also define

$$\rho^{(\epsilon\text{-P})}(\mathscr{C}) \triangleq \max_{l \in [n]} \left\{ \rho^{(\epsilon\text{-P})}(M; \boldsymbol{Q}_l) \right\}$$

as the ϵ-privacy leakage of a given (M, n) IR scheme $\mathscr{C}$.

Note that ϵ-privacy normally gives a *stronger* privacy protection than the MI or MaxL metrics. Moreover, it is worth mentioning that there is a close relation between MaxL and differential privacy [49], [50], see, e.g., [70, Thm. 3]. In this work, although we mainly focus on the MI and MaxL privacy metrics, we will still use the ϵ-privacy metric to show that our proposed Scheme A outperforms the schemes from [60], [61] (see Sections VI and X-B).

Throughout the paper, the information leakage metric of a WPIR scheme $\mathscr{C}$ is denoted by $\rho^{(\cdot)}(\mathscr{C})$, where the superscript indicates the leakage metric (MI, WIL, MaxL, or ϵ-P) we are considering. Moreover, since P_M is fixed, we will also simply write the leakage measure $\rho^{(\cdot)}(\cdot, \cdot)$ as a function of the designed query distribution $P_{\boldsymbol{Q}_l|M}$ of a WPIR scheme. For example, $\rho^{(\mathsf{MI})}(M, \boldsymbol{Q}_l) \equiv \mathsf{I}(P_{\boldsymbol{Q}_l|M}) \equiv \rho^{(\mathsf{MI})}(P_{\boldsymbol{Q}_l|M})$, $l \in [n]$.

D. Download Cost, IR Rate, Upload Cost, and Access Complexity of an (M, n) IR Scheme

For WPIR, in contrast to PIR, the download cost may be different for the retrieval of different files. Thus, the download cost can be defined as the expected download cost over all possible requested files. The download cost of a WPIR scheme $\mathscr{C}$ for the retrieval of the m-th file, denoted by $\mathsf{D}^{(m)}(\mathscr{C})$, is defined as the expected length (in bits) of the returned answers across all servers over all random queries,

$$\mathsf{D}^{(m)}(\mathscr{C}) \triangleq \log_2 |\mathcal{A}| \sum_{l=1}^{n} \mathbb{E}_{\boldsymbol{Q}_l^{(m)}} \left[\ell_l(\boldsymbol{Q}_l^{(m)}) \right],$$

where $\boldsymbol{Q}_l^{(m)}$ is the RV with PMF $P_{\boldsymbol{Q}_l^{(m)}}(\boldsymbol{q}_l) = P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)$. The overall download cost of an IR scheme $\mathscr{C}$, denoted by $\mathsf{D}(\mathscr{C})$, is defined as the expected download cost over all files, i.e.,

$$\mathsf{D}(\mathscr{C}) \triangleq \log_2 |\mathcal{A}| \, \mathbb{E}_M \left[\sum_{l=1}^{n} \mathbb{E}_{\boldsymbol{Q}_l^{(m)}} [\ell_l(\boldsymbol{Q}_l)] \right]$$

$$= \log_2 |\mathcal{A}| \sum_{l=1}^{n} \mathbb{E}_{\boldsymbol{Q}_l} [\ell_l(\boldsymbol{Q}_l)].$$

Accordingly, the IR rate of an IR scheme $\mathscr{C}$ is defined as

$$\mathsf{R}(\mathscr{C}) \triangleq \frac{\beta \log_2 |\mathcal{X}|}{\mathsf{D}(\mathscr{C})}.$$

The upload cost $\mathsf{U}(\mathscr{C})$ of an IR scheme $\mathscr{C}$ is defined as the sum of the entropies of the queries $\boldsymbol{Q}_{[n]}$,

$$\mathsf{U}(\mathscr{C}) \triangleq \sum_{l=1}^{n} \mathsf{H}(\boldsymbol{Q}_l).$$

Moreover, the access complexity $\Delta(\mathscr{C})$ of an IR scheme $\mathscr{C}$ is defined as the expected number of accessed symbols across all servers for the retrieval of a single file,

$$\Delta(\mathscr{C}) \triangleq \sum_{l=1}^{n} \mathbb{E}_{\boldsymbol{Q}_l} [\delta_l(\boldsymbol{Q}_l)] = \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \sum_{l=1}^{n} \mathbb{E}_{\boldsymbol{Q}_l^{(m)}} [\delta_l(\boldsymbol{Q}_l)]. \tag{6}$$

An achievable 4-tuple of an IR scheme is defined as follows.

Definition 2: Consider a DSS with n noncolluding servers storing M files. A 4-tuple $(\mathsf{R}, \mathsf{U}, \Delta, \rho)$ is said to be *achievable* with information leakage metric $\rho^{(\cdot)}$ if there exists an (M, n) IR scheme $\mathscr{C}$ such that $\mathsf{R}(\mathscr{C}) = \mathsf{R}$, $\mathsf{U}(\mathscr{C}) = \mathsf{U}$, $\Delta(\mathscr{C}) = \Delta$, and $\rho^{(\cdot)}(\mathscr{C}) = \rho$. ◇

We remark that a PIR scheme corresponds to an (M, n) IR scheme with $\rho^{(\cdot)} = 0$. It was shown in [7] that for n noncolluding replicated servers and for a given number of files M, the PIR capacity, denoted by $\mathsf{C}_{\mathsf{M}, n}$, is $\mathsf{C}_{\mathsf{M}, n} = \left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{\mathsf{M}-1}}\right)^{-1}$.

III. PARTITION WPIR SCHEME

A simple approach for the construction of WPIR schemes is to first partition the database into η equally-sized partitions, each consisting of $\frac{\mathsf{M}}{\eta}$ files where $\frac{\mathsf{M}}{\eta} \in \mathbb{N}$,[2] and then use a given $\left(\frac{\mathsf{M}}{\eta}, n\right)$ IR scheme to retrieve a file from the corresponding partition. Obviously, the resulting scheme is not a PIR scheme, since the servers gain the knowledge of which partition the requested file belongs to. In this section, we pursue this approach to construct an (M, n) IR scheme building on a given $\left(\frac{\mathsf{M}}{\eta}, n\right)$ IR scheme as a subscheme.

The partition (M, n) WPIR scheme is formally described as follows. Assume the requested file $\boldsymbol{X}^{(m)}$ belongs to the j-th partition, where $j \in [\eta]$. Then, the query $\boldsymbol{Q}_l$ is constructed as

$$\boldsymbol{Q}_l = (\widetilde{\boldsymbol{Q}}_l, j) \in \tilde{\mathcal{Q}}_l \times [\eta], \quad l \in [n], \tag{7}$$

where $\widetilde{\boldsymbol{Q}}_l$ is the query of an existing $\left(\frac{\mathsf{M}}{\eta}, n\right)$ IR scheme $\widetilde{\mathscr{C}}$.

The following theorem states the achievable 4-tuple of the partition scheme.

Theorem 2: Consider a DSS with n noncolluding servers storing M files, and let $\widetilde{\mathscr{C}}$ be an $\left(\frac{\mathsf{M}}{\eta}, n\right)$ IR scheme with achievable 4-tuple $(\tilde{\mathsf{R}}, \tilde{\mathsf{U}}, \tilde{\Delta}, \tilde{\rho}^{(\cdot)})$. Then, the 4-tuple

$$\left(\mathsf{R}(\mathscr{C}), \mathsf{U}(\mathscr{C}), \Delta(\mathscr{C}), \rho^{(\cdot)}(\mathscr{C})\right)$$
$$= \left(\tilde{\mathsf{R}}, \tilde{\mathsf{U}} + n \log_2 \eta, \tilde{\Delta}, \tilde{\rho}^{(\cdot)} + \log_2 \eta\right) \tag{8}$$

is achievable by the (M, n) partition scheme $\mathscr{C}$ constructed from $\widetilde{\mathscr{C}}$ as described in (7). ◇

Proof: The proof is deferred to Appendix A. ∎

---

[2]While it is not necessary that each partition has an equal number of files, for simplicity in this paper we make this assumption.

$P_{\boldsymbol{Q}_1^{(1)}, \boldsymbol{Q}_2^{(1)}}(\boldsymbol{q}_1, \boldsymbol{q}_2)$	$\boldsymbol{q}_2 = (1,0)$	$\boldsymbol{q}_2 = (0,1)$	$P_{\boldsymbol{Q}_1^{(1)}}(\boldsymbol{q}_1)$
$\boldsymbol{q}_1 = (0,0)$	$1-p$	0	$1-p$
$\boldsymbol{q}_1 = (1,1)$	0	p	p
$P_{\boldsymbol{Q}_2^{(1)}}(\boldsymbol{q}_2)$	$1-p$	p	

(a)

$P_{\boldsymbol{Q}_1^{(2)}, \boldsymbol{Q}_2^{(2)}}(\boldsymbol{q}_1, \boldsymbol{q}_2)$	$\boldsymbol{q}_2 = (1,0)$	$\boldsymbol{q}_2 = (0,1)$	$P_{\boldsymbol{Q}_1^{(2)}}(\boldsymbol{q}_1)$
$\boldsymbol{q}_1 = (0,0)$	0	$1-p$	$1-p$
$\boldsymbol{q}_1 = (1,1)$	p	0	p
$P_{\boldsymbol{Q}_2^{(2)}}(\boldsymbol{q}_2)$	p	$1-p$	

(b)

Fig. 1.   The IR rate $\mathsf{R}(\mathscr{C}_\mathsf{A}) \in \left[\frac{2}{3}, 1\right]$ of the proposed $(2,2)$ Scheme A, as a function of $\rho^{(\cdot)}$. The triangle marks the 2-server PIR capacity for $M = 2$.

Since a PIR scheme is also an IR scheme, this simple approach for the construction of WPIR schemes can also be adapted to use any of the existing $\left(\frac{M}{\eta}, n\right)$ PIR schemes in the literature as a subscheme. We refer to the partition scheme that uses a PIR scheme as the underlying subscheme and the query generation in (7) as a *basic scheme* and denote it by $\mathscr{C}^{\mathsf{basic}}$ (it gives the 4-tuple as in (8) with $\tilde{\rho}^{(\cdot)} = 0$). In Section IV-B, we will present another partition WPIR scheme based on our proposed IR scheme.

IV. (M, n) SCHEME A

In [30, Sec. III-B], a PIR scheme that achieves both the minimum upload and download cost was proposed. The queries $\boldsymbol{Q}_{[n]}$ of the scheme in [30, Sec. III-B] are randomly generated according to a random strategy $\boldsymbol{S} = (S_1, \ldots, S_{M-1})$ with i.i.d. entries according to $\mathcal{U}([0 : n - 1])$.[3] In this section, we introduce an (M, n) WPIR scheme, referred to as *Scheme A* and denoted by $\mathscr{C}_\mathsf{A}$, based on the PIR scheme in [30]. Scheme A can be seen as a generalization of the PIR scheme in [30] where we lift the perfect privacy condition (3).

For the proposed scheme, assume that $\mathcal{X} = \{0, 1\}$ and the file size to be $\beta = n - 1$. We represent a query by a length-M vector $\boldsymbol{q}_l = (q_{l,1}, \ldots, q_{l,M}) \in \mathcal{Q}_l \subseteq [0 : n - 1]^M$. Also, the realization of $\boldsymbol{S}$ is denoted by a length-$(M - 1)$ vector $\boldsymbol{s} = (s_1, \ldots, s_{M-1})$, $s_j \in [0 : n - 1]$, $j \in [M - 1]$.

Before describing Scheme A in detail for the general case, for simplicity we first present Scheme A for the case of $M = 2$ files and $n = 2$ servers (i.e., both servers 1 and 2 store $\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$) in the following example.

Example 1: We illustrate the $(2,2)$ Scheme A obtained by adopting a nonuniformly-distributed random strategy $\boldsymbol{S}$ giving a joint PMF $P_{\boldsymbol{Q}_1, \boldsymbol{Q}_2}(\boldsymbol{q}_1, \boldsymbol{q}_2)$ as in Table I. Files $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are composed of one stripe each ($\beta = n - 1 = 1$). The answers $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are given by $\left(\boldsymbol{A}_1(\boldsymbol{q}_1), \boldsymbol{A}_2(\boldsymbol{q}_2)\right) =$

[3]The PIR scheme in [30, Sec. III-B] can be seen as a generalization of the canonical $(2, 2)$ PIR scheme that was first introduced in [27, Sec. III-B] and further elaborated in [72], where the authors focused on the minimization of the storage overhead.
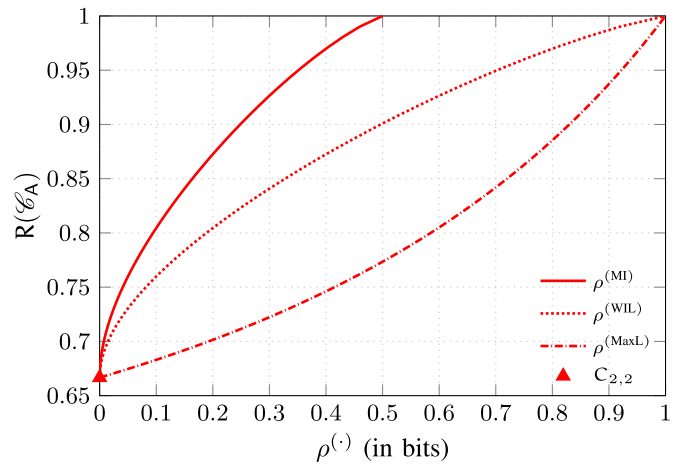
$\left(X_{q_{1,1}}^{(1)} + X_{q_{1,2}}^{(2)}, X_{q_{2,1}}^{(1)} + X_{q_{2,2}}^{(2)}\right)$, where $X_0^{(m)} = 0$ for all $m \in [2]$.

One can easily verify that perfect retrievability is satisfied for the above $(2, 2)$ IR scheme. Its IR rate is a function of p and is given by $\mathsf{R}(p) = (p + (1 - p) + p)^{-1} = (1 + p)^{-1}$. Observe that $M \perp\!\!\!\perp \boldsymbol{Q}_1$, which implies that $\mathsf{I}(M; \boldsymbol{Q}_1) = \mathsf{WIL}(M; \boldsymbol{Q}_1) = \mathsf{MaxL}(M; \boldsymbol{Q}_1) = 0$, i.e., it does not leak any information on the identity of the retrieved file to the first server.

The information leakage is $\rho^{(\mathsf{MI})} = \frac{1 - \mathsf{H}_\mathsf{b}(p)}{2}$, $\rho^{(\mathsf{WIL})} = 1 - \mathsf{H}_\mathsf{b}(p)$, and $\rho^{(\mathsf{MaxL})} = \log_2\left[2(1 - p)\right]$ for $0 \leq p \leq \frac{1}{2}$. From this derivation, it follows that the $(2, 2)$ Scheme A achieves perfect privacy for $p = \frac{1}{2}$. The IR rate of the $(2, 2)$ Scheme A, $\mathsf{R}(\mathscr{C}_\mathsf{A})$, is depicted in Fig. 1 as a function of the information leakage $\rho^{(\cdot)}$. Interestingly, by sacrificing perfect privacy, it is possible to achieve an IR rate larger than the 2-server PIR capacity for 2 files. As expected, the IR rate increases with increasing information leakage. $\Diamond$

Now, we describe Scheme A for the general case of M files and n servers. We assume that the user wants to download file $\boldsymbol{X}^{(m)}$ and has a random strategy $\boldsymbol{S}$ that takes on values $\boldsymbol{s} \in [0 : n - 1]^{M-1}$ with PMF $P_{\boldsymbol{S}}(\boldsymbol{s})$.

1) Query Generation: The query $\boldsymbol{q}_l \in \mathcal{Q}_l$, $l \in [n]$, sent to the l-th server, resulting from the query-encoding function ϕ_l, is defined as

$$\boldsymbol{q}_l = (s_1, \ldots, s_{m-1}, q_{l,m}, s_m, \ldots, s_{M-1}), \qquad (9)$$

where $q_{l,m} \triangleq \left(l - 1 - \sum_{j \in [M-1]} s_j\right) \bmod n$. It follows that

$$\mathcal{Q}_l = \left\{\boldsymbol{q}_l : \left(\sum_{m' \in [M]} q_{l,m'}\right) \bmod n = l - 1\right\}. \qquad (10)$$

Note that the PMF of $\boldsymbol{Q}_l$ conditioned on the file index M satisfies $P_{\boldsymbol{Q}_l^{(m)}}(\boldsymbol{q}_l) = P_{\boldsymbol{S}}(\boldsymbol{s})$.

2) Answer Construction: The answer function φ_l maps the query $\boldsymbol{q}_l$ into

$$\boldsymbol{A}_l = \varphi_l(\boldsymbol{q}_l, \boldsymbol{X}^{[M]}) = X_{q_{l,1}}^{(1)} + \cdots + X_{q_{l,M}}^{(M)}, \qquad (11)$$

where $X_0^{(m')} = 0$ for all $m' \in [M]$. Further, we see that the answer-length functions satisfy

$$\ell_l(\boldsymbol{Q}_l) = \begin{cases} 0 & \text{if } \boldsymbol{q}_l = \boldsymbol{0}, \\ 1 & \text{otherwise.} \end{cases} \quad (12)$$

This completes the construction of the (M, n) Scheme A. Note that it follows from (11) that $\mathcal{A} = \{0, 1\} = \mathcal{X}$. Moreover, using (12), the IR rate of the (M, n) Scheme A, $\mathscr{C}_A$, can be shown to be

$$R(\mathscr{C}_A) = \frac{n-1}{1 - P_{\boldsymbol{Q}_1}(\boldsymbol{0}) + n - 1}. \quad (13)$$

We also remark that if Scheme A uses a random strategy $\boldsymbol{S}$ with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to $\mathcal{U}([0 : n-1])$, then it satisfies (3) and is equivalent to the PIR capacity-achieving scheme proposed in [30].

3) Perfect Retrievability: For completeness, in the following we show that Scheme A satisfies the recovery condition in (1).

From the queries $\boldsymbol{Q}_l$ and answers $\boldsymbol{A}_l$, $l \in [n]$, designed as in (9) and (11), respectively, given $\boldsymbol{S} = \boldsymbol{s} = (s_1, \ldots, s_{M-1}) \in [0 : n-1]^{M-1}$ and $M = m$, the answer from the l-th server can be re-written as

$$\boldsymbol{A}_l = X_{q_{l,m}}^{(m)} + \left(\sum_{m'=1}^{m-1} X_{s_{m'}}^{(m')} + \sum_{m'=m+1}^{M} X_{s_{m'-1}}^{(m')} \right)$$

$$\triangleq X_{q_{l,m}}^{(m)} + Z, \quad l \in [n].$$

Since by definition $q_{l,m} = \left(l - 1 - \sum_{j=1}^{M-1} s_j \right) \bmod n$, $q_{l,m}$ must range thoroughly from 0 to $n-1$, and for $l' - 1 = \left(\sum_{j=1}^{M-1} s_j \right) \bmod n$, we have $\boldsymbol{A}_{l'} = X_{q_{l',m}}^{(m)} + Z = 0 + Z$. Thus, the user can obtain $X_{q_{l,m}}^{(m)} = \boldsymbol{A}_l - Z$, and hence retrieve $\{X_1^{(m)}, \ldots, X_{n-1}^{(m)}\}$.

The following lemma follows immediately from the construction of Scheme A.

Lemma 2: Let $\{S_j\}_{j=1}^{M-1}$ be i.i.d. and $S_j \sim \mathcal{U}([0 : n-1])$ for Scheme A. Then, it satisfies (3) and is equivalent to the PIR capacity-achieving scheme proposed in [30]. $\diamond$

Proof: See Appendix B. ∎

A. $(M, 2)$ Scheme A With $\{S_j\}_{j=1}^{M-1}$ i.i.d. According to Bernoulli (p)

The following result gives an achievable 4-tuple for Scheme A for the case of two servers and a random strategy $\boldsymbol{S} = (S_1, \ldots, S_{M-1})$ with i.i.d. entries according to Bernoulli (p).

Theorem 3: Consider $0 \leq p \leq \frac{1}{2}$. Then, the 4-tuple $(R_A, U_A, \Delta_A, \rho_A^{(\cdot)})$,

$$R_A = \left(1 - (1-p)^{M-1} + 1 \right)^{-1},$$

$$U_A = -\sum_{w=0}^{M} \binom{M}{w} f(w, p) \log_2 f(w, p),$$

$$\Delta_A = \sum_{w=0}^{M} w \binom{M}{w} f(w, p),$$

$$\rho_A^{(MI)} = \frac{U_A}{2} - (M-1) H_b(p),$$

$$\rho_A^{(WIL)} = \log_2 M - \min_{w \in [0:M]} H(M_w), \text{ and}$$

$$\rho_A^{(MaxL)} = \log_2 \sum_{\substack{w \in [M] \\ w : \text{ odd}}} \binom{M}{w} (1-p)^{M-w} p^{w-1}$$

is achievable by the $(M, 2)$ Scheme A with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to Bernoulli (p), where

$$f(w, p) \triangleq \frac{1}{M} \left((M-w)(1-p)^{M-w-1} p^w + w(1-p)^{M-w} p^{w-1} \right)$$

and M_w is a RV with PMF

$$P_{M_w}(m') = \begin{cases} \frac{(1-p)^{M-w-1} p^w}{M f(w,p)} & \text{if } m' \in [M - w], \\ \frac{(1-p)^{M-w} p^{w-1}}{M f(w,p)} & \text{if } m' \in [M - w + 1 : M]. \end{cases} \quad (14)$$

$\diamond$

Proof: See Appendix C. ∎

B. Partition Scheme A: Using Scheme A as a Subscheme

In Section III, we introduced the concept of adopting an existing $\left(\frac{M}{\eta}, n \right)$ IR scheme to retrieve a file from a given partition. In this subsection, unlike (7), where the user sends different queries for different requested files among all partitions, we use a slightly more sophisticated way to construct a WPIR scheme by using Scheme A as a subscheme for every partition. We refer to this scheme as *partition Scheme A* and denote it by $\mathscr{C}_A^{\text{part}}$. In the following, we present the query generation and the answer construction.

1) Query Generation: We consider the j-th partition, $\mathcal{P}_j$, $j \in [\eta]$, containing all files of indices $(j-1)\frac{M}{\eta} + 1, \ldots, j\frac{M}{\eta}$. Given a requested file with index $m = (j-1)\frac{M}{\eta} + m' \in \mathcal{P}_j$, $m' \in \left[\frac{M}{\eta} \right]$, we consider an $\left(\frac{M}{\eta}, n \right)$ Scheme A as a subscheme for partition $\mathcal{P}_j$. The l-th query $\boldsymbol{q}_l \in \mathcal{Q}_l$, $l \in [n]$, is defined as

$$\boldsymbol{q}_l = \Big(\boldsymbol{0}_{1 \times (j-1)\frac{M}{\eta}}, s_1, \ldots, s_{m'-1}, q_{l,(j-1)\frac{M}{\eta} + m'},$$

$$s_{m'}, \ldots, s_{\frac{M}{\eta} - 1}, \boldsymbol{0}_{1 \times (\eta - j)\frac{M}{\eta}} \Big),$$

where $q_{l,(j-1)\frac{M}{\eta} + m'} = \left(l - 1 - \sum_{j \in \left[\frac{M}{\eta} - 1 \right]} s_j \right) \bmod n$. We remark that it is possible that the user sends the all-zero query $\boldsymbol{q}_l = \boldsymbol{0}$ to request different files among all partitions. In this way, since the uncertainty on the requested file is increased, it follows that the leakage of $\mathscr{C}_A^{\text{part}}$ is slightly smaller than the leakage of the basic scheme. Moreover, the query alphabet size is not exactly the same for all servers. In particular, since for every partition, an $\left(\frac{M}{\eta}, n \right)$ Scheme A consists of $n^{\frac{M}{\eta} - 1}$ queries at each server, and only for the first server the all-zero query $\boldsymbol{0}_{1 \times M}$ is sent to retrieve any one of the M files, we have $|\mathcal{Q}_1| = 1 + \eta \left(n^{\frac{M}{\eta} - 1} - 1 \right)$ and $|\mathcal{Q}_l| = \eta \cdot n^{\frac{M}{\eta} - 1}$ for $l \in [2 : n]$.

2) Answer Construction: Similar to Scheme A, the answer function φ_l maps query $\boldsymbol{q}_l$ into $\boldsymbol{A}_l = \varphi_l(\boldsymbol{q}_l, \boldsymbol{X}^{[M]}) = X_{q_{l,1}}^{(1)} + \cdots + X_{q_{l,M}}^{(M)}$, where $X_0^{(m')} = 0$ for all $m' \in [M]$. Further, we see that $\ell_l(\boldsymbol{Q}_l)$ satisfies (12).

3) (M, n) Partition Scheme A With $\{S_j\}_{j=1}^{\frac{\mathsf{M}}{\eta}-1}$ i.i.d. According to $\mathcal{U}([0:n-1])$: We focus on a particular (M, n) partition Scheme A. Since the servers can learn some information from which partition the requested file belongs to, in order to have a relatively small leakage of partition Scheme A, it is reasonable to use Scheme A with $\{S_j\}_{j=1}^{\frac{\mathsf{M}}{\eta}-1}$ i.i.d. according to $\mathcal{U}([0:n-1])$ as a subscheme (i.e., a PIR subscheme, cf. Lemma 2). Thus, this scheme works for an arbitrary number of servers n. We have the following result.

Theorem 4: Let $\frac{\mathsf{M}}{\eta}$ be a positive integer with $\eta \in [\mathsf{M}-1]$. Then, the 4-tuple $\left(\mathsf{R}_{\mathsf{A,P}}, \mathsf{U}_{\mathsf{A,P}}, \Delta_{\mathsf{A,P}}, \rho_{\mathsf{A,P}}^{(\cdot)}\right)$,

$$\mathsf{R}_{\mathsf{A,P}} = \left(1 + \frac{1}{n} + \cdots + \frac{1}{n^{\frac{\mathsf{M}}{\eta}-1}}\right)^{-1},$$

$$\mathsf{U}_{\mathsf{A,P}} = n\left[\left(\frac{\mathsf{M}}{\eta}-1\right)\log_2 n + \log_2 \eta\right] - \frac{\log_2 \eta}{n^{\frac{\mathsf{M}}{\eta}-1}},$$

$$\Delta_{\mathsf{A,P}} = (n-1)\frac{\mathsf{M}}{\eta},$$

$$\rho_{\mathsf{A,P}}^{(\mathsf{MI})} = \log_2 \eta - \frac{\log_2 \eta}{n^{\frac{\mathsf{M}}{\eta}}}, \text{ and}$$

$$\rho_{\mathsf{A,P}}^{(\mathsf{WIL})} = \log_2 \eta = \rho_{\mathsf{A,P}}^{(\mathsf{MaxL})}$$

is achievable by the (M, n) partition Scheme A using the $\left(\frac{\mathsf{M}}{\eta}, n\right)$ Scheme A with $\{S_j\}_{j=1}^{\frac{\mathsf{M}}{\eta}-1}$ i.i.d. according to $\mathcal{U}([0:n-1])$ as a subscheme. ◇

Proof: See Appendix D. ∎

Let $\left(\tilde{\mathsf{R}}, \tilde{\mathsf{U}}, \tilde{\Delta}, 0\right)$ be the achievable 4-tuple of the $\left(\frac{\mathsf{M}}{\eta}, 2\right)$ Scheme A with $\{S_j\}_{j=1}^{\frac{\mathsf{M}}{\eta}-1}$ i.i.d. according to $\mathcal{U}([0:n-1])$. It follows that $\mathsf{U}_{\mathsf{A,P}} = \tilde{\mathsf{U}} + 2\log_2 \eta - \frac{\log_2 \eta}{n^{\frac{\mathsf{M}}{\eta}-1}} < \mathsf{U}(\mathscr{C}^{\mathsf{basic}})$ and $\rho_{\mathsf{A,P}}^{(\mathsf{MI})} = \log_2 \eta - \frac{\log_2 \eta}{n^{\frac{\mathsf{M}}{\eta}}} < \rho^{(\mathsf{MI})}(\mathscr{C}^{\mathsf{basic}})$, while $\mathsf{R}_{\mathsf{A,P}}, \Delta_{\mathsf{A,P}},$ $\rho_{\mathsf{A,P}}^{(\mathsf{WIL})},$ and $\rho_{\mathsf{A,P}}^{(\mathsf{MaxL})}$ are identical to those of the basic scheme $\mathscr{C}^{\mathsf{basic}}$ in Section III (see the details in Appendix D). Hence, in the numerical results section, the results of $\mathscr{C}^{\mathsf{basic}}$ are not presented.

V. CONSTANT-RATE (M, n) SCHEME B

We propose an alternative WPIR scheme, referred to as *Scheme B* and denoted by $\mathscr{C}_{\mathsf{B}}$, based on the PIR scheme in [17, Lem. 4]. Scheme B is constructed as follows. Assume that $\beta = n-1$ and that the user requests file $\boldsymbol{X}^{(m)}$. The random strategy $\boldsymbol{S}$ takes the form of a vector $\boldsymbol{S} = (S_1, \ldots, S_{\beta\mathsf{M}}) \in \mathcal{X}^{\beta\mathsf{M}}$ of length $\beta\mathsf{M}$. The query vector $\boldsymbol{Q}_l \in \mathcal{Q}_l = \mathcal{X}^{\beta\mathsf{M}}$, of length $\beta\mathsf{M}$, is obtained as

$$\boldsymbol{Q}_l = \phi(m, \boldsymbol{S}) = \boldsymbol{S} + \boldsymbol{v}_l^{(m)},$$

where the vector $\boldsymbol{v}_l^{(m)} = (v_{l,1}^{(m)}, \ldots, v_{l,\beta\mathsf{M}}^{(m)})$ is deterministic and is completely determined by $m \in [\mathsf{M}]$. We refer the reader to [17, Sec. V] for details on the design of $\boldsymbol{v}_l^{(m)}$. Briefly, $\boldsymbol{v}_l^{(m)}$ is a binary vector, where $v_{l,i}^{(m)} = 1$ denotes that the i-th symbol is being retrieved from the l-th server. The l-th vector has the following structure,

$$\boldsymbol{v}_l^{(m)} = (\boldsymbol{0}_{1\times(m-1)\beta} \mid \boldsymbol{\Delta}_l \mid \boldsymbol{0}_{1\times(\mathsf{M}-m)\beta}),$$

where $\boldsymbol{\Delta}_l$, $l \in [n-1]$, is the l-th β-dimensional unit vector, and $\boldsymbol{\Delta}_n = \boldsymbol{0}_{1\times\beta}$. The l-th server responds to its corresponding query with the answer $\boldsymbol{A}_l \in \mathcal{A} = \mathcal{X}$ obtained as $\boldsymbol{A}_l = \varphi_l(\boldsymbol{Q}_l, \boldsymbol{X}^{[\mathsf{M}]}) \triangleq \langle \boldsymbol{Q}_l, (X_1^{(1)}, \ldots, X_\beta^{(1)}, X_1^{(2)}, \ldots, X_\beta^{(\mathsf{M})})\rangle$.

For the case where $\{S_j\}_{j=1}^{\beta\mathsf{M}}$ are i.i.d. according to $\mathcal{U}(\mathcal{X})$, Scheme B achieves perfect privacy, and the scheme reduces to the PIR scheme in [17, Lem. 4]. Furthermore, similar to [17, Thm. 2], it can be shown that the scheme achieves perfect retrievability (see (1)), and since its answer-lengths are constant for all possible queries of each server, the IR rate R_{B} of $\mathscr{C}_{\mathsf{B}}$ is equal to $1 - \frac{1}{n}$, irrespective of the information leakage $\rho^{(\cdot)}$.

We remark that for the case of multiple replicated non-colluding servers (uncoded servers), the PIR scheme in [17, Lem. 4] is equivalent to the original PIR protocol proposed in [1], [2]. The designed queries and answers of Scheme B result in a constant rate, while the rate of Scheme A is dependent on the query distribution (see (13)).

In the following subsections, we consider the binary field $\mathcal{X} = \{0, 1\}$.

A. $(\mathsf{M}, 2)$ Scheme B With $\{S_j\}_{j=1}^{\mathsf{M}}$ i.i.d. According to Bernoulli(p)

We have the following result.

Theorem 5: Consider $0 \leq p \leq \frac{1}{2}$. Then, the 4-tuple $\left(\frac{1}{2}, \mathsf{U}_{\mathsf{B}}, \Delta_{\mathsf{B}}, \rho_{\mathsf{B}}^{(\cdot)}\right)$,

$$\mathsf{U}_{\mathsf{B}} = -\sum_{w=0}^{\mathsf{M}} \binom{\mathsf{M}}{w} g(w, p) \log_2 g(w, p) + \mathsf{M}\,\mathsf{H}_{\mathsf{b}}(p),$$

$$\Delta_{\mathsf{B}} = \sum_{w=0}^{\mathsf{M}} w\binom{\mathsf{M}}{w}\left(g(w, p) + h(w, p)\right),$$

$$\rho_{\mathsf{B}}^{(\mathsf{MI})} = \frac{\mathsf{U}_{\mathsf{B}}}{2} - \mathsf{M}\,\mathsf{H}_{\mathsf{b}}(p),$$

$$\rho_{\mathsf{B}}^{(\mathsf{WIL})} = \log_2 \mathsf{M} - \min_{w\in[0:\mathsf{M}]} \mathsf{H}(M_w'), \text{ and}$$

$$\rho_{\mathsf{B}}^{(\mathsf{MaxL})} = \log_2\Bigg((1-p)^{\mathsf{M}-1}p$$
$$+ \sum_{w\in[1:\mathsf{M}]} \binom{\mathsf{M}}{w}(1-p)^{\mathsf{M}-(w-1)}p^{w-1}\Bigg)$$

is achievable by the $(\mathsf{M}, 2)$ Scheme B with $\{S_j\}_{j=1}^{\mathsf{M}}$ i.i.d. according to Bernoulli(p), where $g(w, p) \triangleq \frac{(\mathsf{M}-w)(1-p)^{\mathsf{M}-w-1}p^{w+1}+w(1-p)^{\mathsf{M}-w+1}p^{w-1}}{\mathsf{M}}$, $h(w, p) \triangleq (1-p)^{\mathsf{M}-w}p^w$, and M_w' is a RV with PMF

$$P_{M_w'}(m') = \begin{cases} \frac{(1-p)^{\mathsf{M}-w-1}p^{w+1}}{\mathsf{M}g(w,p)} & \text{if } m' \in [\mathsf{M}-w], \\ \frac{(1-p)^{\mathsf{M}-w+1}p^{w-1}}{\mathsf{M}g(w,p)} & \text{if } m' \in [\mathsf{M}-w+1:\mathsf{M}]. \end{cases}$$

◇

Proof: The proof is similar to the proof of Theorem 3, and is omitted for brevity. ∎

In the following subsection, we analyze the $(\mathsf{M}, 2)$ Scheme B with a uniformly-distributed random strategy $\boldsymbol{S}$. Note that similarly to partition Scheme A in Section IV-B.3, we can also construct a partition scheme by using Scheme B as a subscheme for every partition. We omit the analysis since it

is almost the same as for partition Scheme A, and the result for the (M, n) partition Scheme B with $\{S_j\}_{j=1}^{\frac{M}{n}}$ i.i.d. according to $\mathcal{U}(\mathcal{X})$ is very close to the result in Theorem 4.

B. $(M, 2)$ Scheme B With S Uniformly Distributed

We consider the $(M, 2)$ Scheme B with S uniformly distributed over all length-M binary vectors of weight w. In other words, $S \sim \mathcal{U}(\mathcal{B}_{w,M})$, where $\mathcal{B}_{w,M} \triangleq \{s \in \{0,1\}^M : w_H(s) = w\}$.

Theorem 6: Given any $w \in [0 : M]$. Then, the 4-tuple $\left(\frac{1}{2}, \mathsf{U}_{\mathsf{B,U}}, \Delta_{\mathsf{B,U}}, \rho_{\mathsf{B,U}}^{(\cdot)}\right)$,

$$\mathsf{U}_{\mathsf{B,U}} = \log_2 \binom{M}{w} + y(w, M),$$

$$\Delta_{\mathsf{B,U}} = 1 + 2w\left(1 - \frac{1}{M}\right),$$

$$\rho_{\mathsf{B,U}}^{(\mathsf{MI})} = \frac{y(w, M) - \log_2 \binom{M}{w}}{2},$$

$$\rho_{\mathsf{B,U}}^{(\mathsf{WIL})} = \log_2 M$$
$$\qquad - \min\{\log_2(w+1), \log_2(M - w + 1)\}, \text{ and}$$

$$\rho_{\mathsf{B,U}}^{(\mathsf{MaxL})} = \log_2\left(\frac{M - w}{w + 1} + \frac{w}{M - w + 1}\right)$$

is achievable by the $(M, 2)$ Scheme B with $S \sim \mathcal{U}(\mathcal{B}_{w,M})$, where $y(w, M) \triangleq \log_2\binom{M}{w} + \log_2 M - \frac{(M-w)\log_2(w+1)}{M} - \frac{w\log_2(M-w+1)}{M}$. ◇

Proof: See Appendix E. ∎

We remark that the analysis of the $(M, 2)$ Scheme A with $S \sim \mathcal{U}(\mathcal{B}_{w,M-1})$ can also be done by following the same approach as for Theorem 6.[4] However, since the resulting performance is much worse than those of the aforementioned WPIR schemes for the case of $n = 2$ servers, we omit the detailed analysis in this paper.

In the rest of the paper, except for the next Section VI and the numerical results in Section X, we consider only the MI and MaxL privacy metrics as these are more commonly used [56].

VI. ϵ-PRIVACY FOR SCHEME A

In this section, we focus on the leakage analysis of the (M, n) Scheme A under the ϵ-privacy metric. Since the WPIR rate of the (M, n) Scheme A is equal to (13), irrespective of the used privacy metric, we only need to focus on the design of the conditional query distribution $P_{Q_l^{(m)}}(q_l)$ at each server.

To compare our results with the works in [60] and [61], we first summarize the achievable rate for a given leakage constraint $\rho \geq 0$ as follows.

Lemma 3 ([61, Eq. (26)]): Consider a DSS with n non-colluding servers storing M files. Then, given an ϵ-privacy leakage constraint $\rho^{(\epsilon\text{-P})} \leq \rho$ with $\rho \geq 0$, the rate

$$\mathsf{R}_{\mathsf{LPIR}}^{(\epsilon\text{-P})} = \left(1 + \frac{n^{M-1} - 1}{(n-1)(e^\rho + n^{M-1} - 1)}\right)^{-1} \quad (15)$$

is achievable. Moreover, the WPIR rate under the ϵ-privacy metric is bounded from above by

$$\mathsf{R}^{(\epsilon\text{-P})}(\mathscr{C}) \leq \frac{1 - \frac{1}{ne^\rho}}{1 - \frac{1}{(ne^\rho)^M}} \triangleq \mathsf{R}_{\mathsf{UB}}^{(\epsilon\text{-P})}. \quad (16)$$

◇

In [60] and [61], the authors proposed the *path-based approach* across databases to obtain the achievable download rate $\mathsf{R}_{\mathsf{LPIR}}^{(\epsilon\text{-P})}$ in (15). In the following subsection, we show that it is possible to achieve a better tradeoff between the ϵ-privacy leakage and the download rate by using Scheme A.

A. $(M, 2)$ Scheme A With $\{S_j\}_{j=1}^{M-1}$ i.i.d. According to a Bernoulli-Distributed RV

Theorem 7: Consider $0 < p_\rho \triangleq (1 + e^\rho)^{-1} \leq \frac{1}{2}$ for $\rho \geq 0$. Then, given an ϵ-privacy leakage constraint $\rho^{(\epsilon\text{-P})} \leq \rho$, the rate

$$\mathsf{R}_{\mathsf{A}}^{(\epsilon\text{-P})} = \left(1 - (1 - p_\rho)^{M-1} + 1\right)^{-1} \quad (17)$$

is achievable by the $(M, 2)$ Scheme A with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to Bernoulli (p_ρ). ◇

Proof: See Appendix F. ∎

VII. MINIMIZATION OF THE INFORMATION LEAKAGE FOR SCHEME A WITH TIME-SHARING

A main objective of this work is to determine the optimal WPIR scheme that leaks the smallest amount of information, subject to a given IR download cost, upload cost, or access complexity. Since both the information leakage and the IR rate can be improved based on the query generation of Scheme A, our aim is to study the optimal tradeoff between the information leakage and the download cost for Scheme A. In the rest of paper, we will mainly focus on the MI and MaxL privacy metrics.

Following the notion of Theorem 1 for the MI metric, we can use the time-sharing principle to construct a time-sharing scheme from Scheme A. In particular, consider the Scheme A $\mathscr{C}_A$ with query-encoding functions ϕ_l, answer functions φ_l, and a random strategy S presented in Section IV. We design the query-encoding functions $\bar{\phi}_l = \phi_{\sigma^{T-1}(l)}(M, S)$ and the answer functions $\bar{\varphi}_l = \varphi_{\sigma^{T-1}(l)}\left(\phi_{\sigma^{T-1}(l)}(M, S), X^{[M]}\right)$ of a given requested file index M to construct the time-sharing scheme of Scheme A, where $T \sim \mathcal{U}([n])$. Such a scheme is referred to as time-sharing Scheme A and denoted by $\overline{\mathscr{C}}_A$.[5] Recall that in Scheme A the query realization q_l of the l-th server, $l \in [n]$, belongs to $\mathcal{Q}_l$, defined in (10), from which it follows that all of the query sets $\mathcal{Q}_l$ are distinct. The conditional query PMF $P_{Q_l^{(m)}}(q_l) = P_S(s)$ is independent of $M = m$, and the download cost of the (M, n) Scheme A is $1 - P_S(0) + (n - 1)$ (cf. Section IV). Denote by $z_s \triangleq P_S(s)$ the PMF of the random strategy S, and $q_l \setminus \{m\} \triangleq (q_{l,1}, \ldots, q_{l,m-1}, q_{l,m+1}, \ldots, q_{l,M})$, $m \in [M]$. By applying the time-sharing approach, the resulting query set at the l-th

---

[4] The scheme is not equal to that of [30] because of the difference in the vector space of the random strategy. The former involves all length-$(M-1)$ vectors of weight w, while the latter consists of all vectors of length $M - 1$.

[5] The time-sharing principle can be applied to any WPIR scheme. However, here we concentrate only on the time-sharing Scheme A.

server is $\overline{\mathcal{Q}}_l = [0 : n-1]^M$, and the conditional PMF of $\overline{\boldsymbol{Q}}_l$ given $M = m$ is

$$
P_{\overline{\boldsymbol{Q}}_l^{(m)}}(\bar{\boldsymbol{q}}_l) = \sum_{t=1}^n \Pr[T = t] \Pr\left[\boldsymbol{Q}_{\sigma^{t-1}(l)}^{(m)} = \bar{\boldsymbol{q}}_l \Big| T = t\right]
$$
$$
= \frac{1}{n} z_{\boldsymbol{s}}, \quad \text{for } \boldsymbol{s} = \bar{\boldsymbol{q}}_l \setminus \{m\}, \ \bar{\boldsymbol{q}}_l \in \overline{\mathcal{Q}}_l, \qquad (18)
$$

where (18) follows since all query sets in Scheme A are different. In other words, by using the time-sharing approach, we obtain a new scheme with a conditional query distribution at each server equal to

$$
\left(\frac{P_{\boldsymbol{Q}_l^{(m)}_{\sigma^0(l)}}(\cdot)}{n}, \cdots, \frac{P_{\boldsymbol{Q}_l^{(m)}_{\sigma^{n-1}(l)}}(\cdot)}{n}\right), \qquad (19)
$$

where $P_{\boldsymbol{Q}_l^{(m)}}(\cdot) = (p_1, \ldots, p_{|\mathcal{Q}_l|})$ represents the conditional query distribution corresponding to $P_{Q_l|M=m}$ for Scheme A. Therefore, from (18) or (19) it follows that every server has identical information leakage for the time-sharing Scheme A under both the MI and MaxL privacy metrics. Note that the download cost stays the same as for Scheme A.

For the time-sharing Scheme A, the minimization of the information leakage $\rho^{(\cdot)}(\overline{\mathscr{C}}_A)$ under a download cost constraint D can be cast as the optimization problem

$$
\begin{aligned}
\text{minimize} \quad & \rho^{(\cdot)}(\overline{\mathscr{C}}_A) && (20a)\\
\text{subject to} \quad & 1 - z_{\boldsymbol{0}} + (n-1) \leq D, && (20b)\\
& \sum_{\boldsymbol{s} \in [0:n-1]^{M-1}} z_{\boldsymbol{s}} = 1. && (20c)
\end{aligned}
$$

A. Optimizing the MI Leakage

In terms of the MI privacy metric, we know from Theorem 1 that the leakage of $\overline{\mathscr{C}}_A$ is equal to[6]

$$
\rho^{(\mathsf{MI})}(\overline{\mathscr{C}}_A) = \mathsf{I}(M; \overline{\boldsymbol{Q}}_l)
$$
$$
= \frac{1}{n} \sum_{l' \in [n]} \left[\mathsf{H}(\boldsymbol{Q}_{l'}) - \mathsf{H}(\boldsymbol{Q}_{l'}|M)\right]
$$
$$
= \frac{1}{n} \sum_{l' \in [n]} \mathsf{H}(P_{\boldsymbol{Q}_{l'}}) - \mathsf{H}\left((z)_{\boldsymbol{s} \in [0:n-1]^{M-1}}\right)
$$
$$
\overset{(a)}{=} \frac{1}{n} \sum_{\bar{\boldsymbol{q}}_l \in [0:n-1]^M} \left[\left(-\frac{\sum_{m=1}^M P_{\boldsymbol{S}}(\bar{\boldsymbol{q}}_l \setminus \{m\})}{M}\right)\right.
$$
$$
\left. \cdot \log_2 \left(\frac{\sum_{m=1}^M P_{\boldsymbol{S}}(\bar{\boldsymbol{q}}_l \setminus \{m\})}{M}\right)\right] - \mathsf{H}\left((z)_{\boldsymbol{s} \in [0:n-1]^{M-1}}\right)
$$
$$
= \frac{1}{n} \sum_{\bar{\boldsymbol{q}}_1 \in [0:n-1]^M} \left(-\frac{\sum_{m=1}^M z_{\bar{\boldsymbol{q}}_1 \setminus \{m\}}}{M}\right) \log_2 \left(\frac{\sum_{m=1}^M z_{\bar{\boldsymbol{q}}_1 \setminus \{m\}}}{M}\right)
$$
$$
- \mathsf{H}\left((z)_{\boldsymbol{s} \in [0:n-1]^{M-1}}\right), \qquad (21)
$$

where (a) holds by the definition of entropy and the fact that $\bigcup_{l=1}^n \mathcal{Q}_l = \overline{\mathcal{Q}}_1 = [0:n-1]^M$. Hence, (20a) becomes (21).

---

[6]Note that the MI leakage of Scheme A and that of the corresponding time-sharing Scheme A is always the same due to the definition of MI leakage in (4).

We remark that the MI objective function $\mathsf{I}(M; \overline{\boldsymbol{Q}}_1)$ is convex in $P_{\overline{\boldsymbol{Q}}_1|M}$, and $P_{\overline{\boldsymbol{Q}}_1|M}$ is subject to the following linear constraints

$$
P_{\overline{\boldsymbol{Q}}_1|M}(\bar{\boldsymbol{q}}_1|m) = P_{\overline{\boldsymbol{Q}}_1|M}(\bar{\boldsymbol{q}}_1'|m), \ \forall \bar{\boldsymbol{q}}_1 \setminus \{m\} = \bar{\boldsymbol{q}}_1' \setminus \{m\}.
$$

Thus, the optimization problem (20) under MI leakage is convex. However, it is difficult to have closed-form optimal solutions for $(M, n) \neq (2, 2)$, and hence instead we present numerical results of the optimized time-sharing $(M, 2)$ Scheme A for several values of the number of files M in Section X. Lastly, we would like to emphasize that for any information leakage metric $\rho^{(\cdot)}$ that is convex in $P_{\overline{\boldsymbol{Q}}_1|M}$ we end up with a convex optimization problem for the maximization of the download rate of the time-sharing Scheme A.

B. Optimizing the MaxL

In this subsection, we turn our attention to the minimization of the MaxL for the proposed (M, n) Scheme A with time-sharing. Similar to the derivation for the MI metric, by definition (20a) becomes

$$
\rho^{(\mathsf{MaxL})}(\overline{\mathscr{C}}_A) = \log_2 \sum_{\bar{\boldsymbol{q}}_l \in \overline{\mathcal{Q}}_l} \max_{m \in [M]} \frac{z_{\bar{\boldsymbol{q}}_l \setminus \{m\}}}{n}.
$$

We remark that as for MI leakage, the time-sharing Scheme A $\overline{\mathscr{C}}_A$ also has identical MaxL at each server. Using again the fact that the sets $\mathcal{Q}_l$ are distinct and $\bigcup_{l=1}^n \mathcal{Q}_l = \overline{\mathcal{Q}}_l$, we have

$$
\rho^{(\mathsf{MaxL})}(\mathscr{C}_A) = \max_{l \in [n]} \left\{\log_2 \sum_{\boldsymbol{q}_l \in \mathcal{Q}_l} \max_{m \in [M]} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)\right\}
$$
$$
= \log_2 \left\{\max_{l \in [n]} \sum_{\boldsymbol{q}_l \in \mathcal{Q}_l} \max_{m \in [M]} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)\right\}
$$
$$
\geq \log_2 \left\{\frac{1}{n} \sum_{l \in [n]} \sum_{\boldsymbol{q}_l \in \mathcal{Q}_l} \max_{m \in [M]} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)\right\}
$$
$$
= \rho^{(\mathsf{MaxL})}(\overline{\mathscr{C}}_A).
$$

From Lemma 1 and using a similar argument as in Section VII-A, replacing the objective function with $2^{\rho^{(\mathsf{MaxL})}(\overline{\mathscr{C}}_A)}$ in (20) gives a convex minimization problem. In Section X below, we give numerical optimal values for (20) under MaxL and compare them with the converse results presented next.

We remark here that the minimization of the information leakage for Scheme A with time-sharing under a download cost constraint D can also be done for the ϵ-privacy metric. However, as this work mainly focuses on the MI and MaxL privacy metrics, we leave the analysis of this minimization of the ϵ-privacy metric as future work.

VIII. CONVERSE RESULTS FOR MI LEAKAGE

In order to present the converse results of WPIR for the MI metric, we first introduce the following measure between two PMFs.

Definition 3: The TV distance between two PMFs P_{Y_1} and P_{Y_2} on the same finite alphabet $\mathcal{Y}$ is defined as

$$
\|P_{Y_1} - P_{Y_2}\|_{\mathsf{TV}} \triangleq \max_{\mathcal{Z} \subseteq \mathcal{Y}} |P_{Y_1}(\mathcal{Z}) - P_{Y_2}(\mathcal{Z})|,
$$

where $P_Y(\mathcal{Z}) \triangleq \sum_{z \in \mathcal{Z}} P_Y(z)$ is the probability of all realizations in the set $\mathcal{Z}$. $\diamond$

Next, we review a useful lemma related to the TV distance, which was presented in [54, Lem. 2].

Lemma 4: If $\mathsf{I}(X;Y) \leq \rho$, then for any $x, x' \in \mathcal{X}$, we have

$$\left\| P_{Y|X=x} - P_{Y|X=x'} \right\|_{\mathsf{TV}} \leq 1 - 2\,\mathsf{H}_{\mathsf{b}}^{-1}(1-\rho) \triangleq \delta^{\mathsf{MI}}. \quad \diamond$$

Here, since we require that $\delta^{\mathsf{MI}} \geq 0$, it is easy to see that we can specify $\mathsf{H}_{\mathsf{b}}^{-1}(1-\rho) \in \left[0, \frac{1}{2}\right]$.

Lastly, we consider a known result between the entropy difference and the TV distance, which can be derived by using a probabilistic coupling technique.

Lemma 5 ([73, Eq. (4)]): If $\left\| P_{Y|X=x} - P_{Y|X=x'} \right\|_{\mathsf{TV}} \leq \delta^{\mathsf{MI}}$, then

$$\begin{aligned}
\left| \mathsf{H}\left(P_{Y|X=x}\right) - \mathsf{H}\left(P_{Y|X=x'}\right) \right| & \\
\leq \delta^{\mathsf{MI}} \log_2(|\mathcal{Y}|-1) + \mathsf{H}_{\mathsf{b}}(\delta^{\mathsf{MI}}) & \quad (22) \\
\leq \delta^{\mathsf{MI}} \log_2(|\mathcal{Y}|-1) + 1. & \quad \diamond
\end{aligned}$$

Note that in the inequality (23), the upper bound becomes tight for $0 \leq \delta^{\mathsf{MI}} \leq 1 - \frac{1}{|\mathcal{Y}|}$ [74, Thm. 6].[7]

We remark that under the MI metric, we measure the overall leakage of a WPIR scheme in an average manner, i.e., $\rho^{(\mathsf{MI})} = \frac{1}{n} \sum_{l \in [n]} \mathsf{I}(M; \boldsymbol{Q}_l)$. Here, given a leakage constraint $\rho^{(\mathsf{MI})} \leq \rho$, we assume that $\mathsf{I}(M; \boldsymbol{Q}_l) = \rho_l, \forall l \in [n]$.

Let us define

$$\begin{aligned}
\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) &\triangleq \delta_l^{\mathsf{MI}} \log_2(|\mathcal{Q}_l \times \mathcal{A}_l| - 1) + 1, \\
\epsilon^{\mathsf{MI}}(\mathcal{Q}_l) &\triangleq \delta_l^{\mathsf{MI}} \log_2(|\mathcal{Q}_l| - 1) + 1,
\end{aligned}$$

where $\delta_l^{\mathsf{MI}} = 1 - 2\,\mathsf{H}_{\mathsf{b}}^{-1}(1-\rho_l)$ and $\mathcal{A}_l$ is the alphabet of the answer $\boldsymbol{A}_l, l \in [n]$. We give the following useful lemma.

Lemma 6: Given $m \neq m'$, where $m, m' \notin \mathcal{M} \subsetneq [\mathsf{M}-1]$, we have

$$\begin{aligned}
\mathsf{H}(\boldsymbol{A}_{[n]}^{(m)} | \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}) & \\
\geq \beta \log_2 |\mathcal{X}| + \frac{\mathsf{H}\left(\boldsymbol{A}_{[n]}^{(m')} | \boldsymbol{Q}_{[n]}^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\right)}{n} & \\
- \frac{\sum_{l=1}^{n} \left[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\right]}{n}. & \quad (23)
\end{aligned}$$

Moreover,

$$\mathsf{H}(\boldsymbol{A}_{[n]}^{(\mathsf{M})} | \boldsymbol{Q}_{[n]}^{(\mathsf{M})}, \boldsymbol{X}^{[\mathsf{M}-1]}) \geq \beta \log_2 |\mathcal{X}|. \quad (24)$$

$\diamond$

Proof: The proof is deferred to Appendix G. ∎

Now, we are ready to derive a general lower bound on D. Since $\mathsf{H}(\boldsymbol{A}_l^{(m)} | \boldsymbol{Q}_l^{(m)} = \boldsymbol{q}_l) \leq \log_2 |\mathcal{A}|^{\ell_l(\boldsymbol{q}_l)}$ for a given

$\boldsymbol{q}_l \in \mathcal{Q}_l$, we have

$$\begin{aligned}
\mathsf{D}^{(m)} &= \log_2 |\mathcal{A}| \sum_{l=1}^{n} \sum_{\boldsymbol{q}_l \in \mathcal{Q}_l} P_{\boldsymbol{Q}_l^{(m)}}(\boldsymbol{q}_l) \ell_l(\boldsymbol{q}_l) \\
&\geq \sum_{l=1}^{n} \sum_{\boldsymbol{q}_l \in \mathcal{Q}_l} P_{\boldsymbol{Q}_l^{(m)}}(\boldsymbol{q}_l) \mathsf{H}\left(\boldsymbol{A}_l^{(m)} | \boldsymbol{Q}_l^{(m)} = \boldsymbol{q}_l\right) \\
&= \sum_{l=1}^{n} \mathsf{H}(\boldsymbol{A}_l^{(m)} | \boldsymbol{Q}_l^{(m)}).
\end{aligned}$$

Note that without loss of generality, we can assume that the conditional entropies $\mathsf{H}(\boldsymbol{A}_{[n]}^{(m)} | \boldsymbol{Q}_{[n]}^{(m)})$, $m \in [\mathsf{M}]$, satisfy $\mathsf{H}(\boldsymbol{A}_{[n]}^{(1)} | \boldsymbol{Q}_{[n]}^{(1)}) \leq \cdots \leq \mathsf{H}(\boldsymbol{A}_{[n]}^{(\mathsf{M})} | \boldsymbol{Q}_{[n]}^{(\mathsf{M})})$.

Hence, similar to the recursive procedure used in [7, Sec. V-A], [12, Sec. VI], the total download cost can be bounded from below by

$$\begin{aligned}
\mathsf{D}(\mathscr{C}) & \\
= \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \mathsf{D}^{(m)} &\geq \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \sum_{l=1}^{n} \mathsf{H}\left(\boldsymbol{A}_l^{(m)} | \boldsymbol{Q}_l^{(m)}\right) \\
&\stackrel{(a)}{\geq} \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \mathsf{H}(\boldsymbol{A}_{[n]}^{(m)} | \boldsymbol{Q}_{[n]}^{(m)}) \\
&\geq \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \mathsf{H}(\boldsymbol{A}_{[n]}^{(1)} | \boldsymbol{Q}_{[n]}^{(1)}) = \mathsf{H}(\boldsymbol{A}_{[n]}^{(1)} | \boldsymbol{Q}_{[n]}^{(1)}) \\
&\stackrel{(b)}{\geq} \beta \log_2 |\mathcal{X}| + \frac{\mathsf{H}\left(\boldsymbol{A}_{[n]}^{(2)} | \boldsymbol{Q}_{[n]}^{(2)}, \boldsymbol{X}^{(1)}\right)}{n} \\
&\quad - \frac{\sum_{l=1}^{n} \left[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\right]}{n} \\
&\vdots \\
&\stackrel{(c)}{\geq} \beta \log_2 |\mathcal{X}| + \sum_{m=1}^{\mathsf{M}-1} \left[\frac{\beta \log_2 |\mathcal{X}|}{n^m} \right. \\
&\quad \left. - \frac{\sum_{l=1}^{n} \left[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\right]}{n^m} \right],
\end{aligned}$$
(25)

where (a) holds because conditioning reduces entropy, and (b)–(c) follow by recursively applying Lemma 6 M times with $\mathcal{M} = \emptyset, \{1\}, \ldots, [\mathsf{M}-1]$, respectively. It is worth mentioning that under the assumption of perfect privacy, i.e., $\delta_l^{\mathsf{MI}} = 0$, $\forall l \in [n]$, $\mathsf{H}(\boldsymbol{A}_{[n]}^{(m)} | \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}})$ in (23) can be bounded from below by

$$\begin{aligned}
\mathsf{H}(\boldsymbol{A}_{[n]}^{(m)} | \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}) & \\
\geq \beta \log_2 |\mathcal{X}| + \frac{\mathsf{H}\left(\boldsymbol{A}_{[n]}^{(m')} | \boldsymbol{Q}_{[n]}^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\right)}{n}
\end{aligned}$$

by bounding (38) using (23) directly in the proof of Lemma 6 (see Appendix G). Following the above recursive steps, it can then be shown that the same converse results of the PIR capacity proof in [7, Sec. V-A] can be obtained.

---

[7]The results shown in [73] and [74] assume the variational distance as the measure between two PMFs. It can be easily shown that the TV distance is equal to the variational distance divided by 2, i.e., $\left\| P_{Y_1} - P_{Y_2} \right\|_{\mathsf{TV}} = \frac{1}{2} \sum_{y \in \mathcal{Y}} \left| P_{Y_1}(y) - P_{Y_2}(y) \right|$, see, e.g., [75, Lem. 3.12].

A. A Converse Bound With Restricted Alphabets of Queries and Answers

The expression in (25) indicates that a general lower bound on D can be arbitrarily dependent on the choice of the alphabets of the queries and answers. In this subsection, a converse bound for WPIR schemes is derived from a practical design perspective. Note that in the information-theoretic PIR setup, the upload cost can be ignored as it does not scale with the file size. Moreover, to have an efficient WPIR scheme, the downloaded answer size per server should be smaller than or equal to the entire retrieved file size. Hence, we assume that

$$|\mathcal{Q}_l| \le \alpha < \infty, \tag{26a}$$

$$|\mathcal{A}_l| \le |\mathcal{X}|^\beta, \tag{26b}$$

$\forall l \in [n]$, for some positive $\alpha \in \mathbb{N}$, i.e., the query sizes are finite and each answer takes value on a smaller alphabet than that of the retrieved file.

We first state the following theorem, which gives an upper bound on the maximum possible WPIR rate for the MI metric.

Theorem 8: Consider an (M, n) WPIR scheme that satisfies (26) and with MI leakage $\rho^{(\text{MI})} \le \rho$. Then, the maximum possible WPIR rate, denoted by $\mathsf{R}_{\text{max}}^{(\text{MI})}$, is bounded from above by

$$\mathsf{R}_{\text{max}}^{(\text{MI})} \le \left[\frac{1}{n^{M-1}} + 2 \sum_{m=1}^{M-1} \frac{1}{n^{m-1}} \mathsf{H}_{\mathsf{b}}^{-1}(1-\rho) \right]^{-1} \triangleq \mathsf{R}_{\text{UB}}^{(\text{MI})}. \quad \diamond$$

Proof: Under these assumptions, (25) becomes

$$\mathsf{D}(\mathscr{C})$$

$$\ge \beta \log_2 |\mathcal{X}| + \sum_{m=1}^{M-1} \left[\frac{\beta \log_2 |\mathcal{X}|}{n^m} - \sum_{l=1}^{n} \left(\frac{\delta_l^{\text{MI}} \log_2 (\alpha |\mathcal{X}|^\beta)}{n^m} \right. \right.$$
$$\left. \left. + \frac{\delta_l^{\text{MI}} \log_2 \alpha}{n^m} + \frac{2}{n^m} \right) \right]. \tag{27}$$

Dividing (27) by $\beta \log_2 |\mathcal{X}|$ gives

$$\frac{\mathsf{D}(\mathscr{C})}{\beta \log_2 |\mathcal{X}|}$$

$$\ge 1 + \sum_{m=1}^{M-1} \left[\frac{1}{n^m} - \sum_{l=1}^{n} \left(\frac{\delta_l^{\text{MI}} \log_2 \alpha}{n^m \beta \log_2 |\mathcal{X}|} + \frac{\delta_l^{\text{MI}}}{n^m} \right. \right.$$
$$\left. \left. + \frac{\delta_l^{\text{MI}} \log_2 \alpha}{n^m \beta \log_2 |\mathcal{X}|} + \frac{2}{n^m \beta \log_2 |\mathcal{X}|} \right) \right] \tag{28}$$

$$\xrightarrow{\beta \to \infty} 1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{\sum_{l=1}^{n} \delta_l^{\text{MI}}}{n^m} \tag{29}$$

$$= 1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{\sum_{l=1}^{n} \left[1 - 2 \mathsf{H}_{\mathsf{b}}^{-1}(1-\rho_l) \right]}{n^m}$$

$$= 1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{n}{n^m} + 2 \sum_{m=1}^{M-1} \sum_{l=1}^{n} \frac{\mathsf{H}_{\mathsf{b}}^{-1}(1-\rho_l)}{n^m}$$

$$\overset{(a)}{\ge} \frac{1}{n^{M-1}} + 2 \sum_{m=1}^{M-1} \frac{n}{n^m} \mathsf{H}_{\mathsf{b}}^{-1} \left(1 - \frac{1}{n} \sum_{l=1}^{n} \rho_l \right)$$

$$\overset{(b)}{\ge} \frac{1}{n^{M-1}} + 2 \sum_{m=1}^{M-1} \frac{n}{n^m} \mathsf{H}_{\mathsf{b}}^{-1}(1-\rho)$$

$$= \frac{1}{n^{M-1}} + 2 \sum_{m=1}^{M-1} \frac{1}{n^{m-1}} \mathsf{H}_{\mathsf{b}}^{-1}(1-\rho),$$

where (a) and (b) hold because the inverse binary entropy function is convex and increasing in $[0, 1]$, respectively. ∎

We remark that in the proof of Theorem 8 (as well as in the following converse results), the subtlety is to allow the file size β to go to infinity, such that we can have nontrivial converse bounds, i.e., $\mathsf{R}_{\text{UB}}^{(\cdot)} \le 1$, for certain cases of (M, n). We will also discuss the tightness of our proposed converse results in Section X-C. Note that the converse bound in (28) is in general trivial when β decreases.

In the following, we prove the largest possible achievable WPIR rate for the special case of $(M, n) = (2, 2)$ under the additional constraint that only one of the two servers can leak. $\diamond$

Theorem 9: Consider an $(M, n) = (2, 2)$ WPIR scheme that satisfies (26) and with MI leakage $\rho^{(\text{MI})} \le \rho \le 1$. Then, the maximum possible WPIR rate is

$$\mathsf{R}_{\text{max}}^{(\text{MI})}(\rho) = \left[1 + \mathsf{H}_{\mathsf{b}}^{-1}(1-2\rho) \right]^{-1}$$

under the assumption that only one of the two servers can leak. $\diamond$

Proof: The achievable scheme is the Scheme A presented in Example 1. Thus, we only need to prove the converse. Assume without loss of generality that $\mathsf{I}(M; \boldsymbol{Q}_1) = \rho_1 = 0$ and $\mathsf{I}(M; \boldsymbol{Q}_2) = \rho_2 = 2\rho$. By definition, $\delta_1^{\text{MI}} = 1 - 2 \mathsf{H}_{\mathsf{b}}^{-1}(1 - \rho_1) = 0$ and we can use the exact same derivation as in Section VIII-A to obtain from (30)

$$\frac{\mathsf{D}(\mathscr{C})}{\beta \log_2 |\mathcal{X}|} \ge 1 + \frac{1}{2} - \frac{\delta_2^{\text{MI}}}{2}$$

$$= \frac{3}{2} - \frac{1 - 2 \mathsf{H}_{\mathsf{b}}^{-1}(1-2\rho)}{2} = 1 + \mathsf{H}_{\mathsf{b}}^{-1}(1-2\rho),$$

as $\beta \to \infty$, which completes the proof. ∎

IX. CONVERSE RESULTS FOR MAXL

In this section, we present the converse results for the MaxL metric. Similar to the case of the MI privacy metric, see Theorem 10 presented here.

Theorem 10: Consider an (M, n) WPIR scheme that satisfies (26) and with MaxL $\rho^{(\text{MaxL})} \le \rho$. Then, the maximum possible WPIR rate, denoted by $\mathsf{R}_{\text{max}}^{(\text{MaxL})}$, is bounded from above by

$$\mathsf{R}_{\text{max}}^{(\text{MaxL})} \le \left[1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{2^\rho - 1}{n^{m-1}} \right]^{-1} \triangleq \mathsf{R}_{\text{UB}}^{(\text{MaxL})}. \quad \diamond$$

Proof: Similar to the case with the MI privacy metric, we first make use of the following lemma.

Lemma 7: If $\mathsf{MaxL}(X; Y) \le \rho$, then for any $x, x' \in \mathcal{X}$, we have

$$\left\| P_{Y|X=x} - P_{Y|X=x'} \right\|_{\text{TV}} \le 2^\rho - 1 \triangleq \delta^{\text{MaxL}}. \quad \diamond$$

Lemma 7 can be proven by a similar argument as in [54, App. C]. The proof is provided in Appendix H for completeness.

Note that for the MaxL metric, we consider the worst-case MaxL over all servers, i.e., $\rho^{(\text{MaxL})} = \max_{l \in [n]} \text{MaxL}(M; \boldsymbol{Q}_l)$. If the leakage at the l-th server is $\text{MaxL}(M; \boldsymbol{Q}_l) \leq \rho_l$, then we have $\rho^{(\text{MaxL})} \leq \max_{l \in [n]} \rho_l \triangleq \rho$. A lower bound on the download cost can be proven by following the same steps as in Section VIII. Under the assumptions in (26), and as $\beta \to \infty$, (30) becomes

$$\frac{\text{D}(\mathscr{C})}{\beta \log_2 |\mathcal{X}|} \geq 1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{\sum_{l=1}^{n} \delta_l^{\text{MaxL}}}{n^m}$$

$$\geq 1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{n \max_{l \in [n]} \delta_l^{\text{MaxL}}}{n^m}$$

$$= 1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{\max_{l \in [n]} \delta_l^{\text{MaxL}}}{n^{m-1}}$$

$$= 1 + \sum_{m=1}^{M-1} \frac{1}{n^m} - \sum_{m=1}^{M-1} \frac{2^\rho - 1}{n^{m-1}}, \tag{30}$$

where (30) follows from Lemma 7 and $\max_{l \in [n]} \delta_l^{\text{MaxL}} = 2^{\max_{l \in [n]} \rho_l} - 1 = 2^\rho - 1$. ∎

In the following theorem we give the maximum achievable WPIR rate under the MaxL metric for the special case of $(M, n) = (2, 2)$.

Theorem 11: Consider an $(M, n) = (2, 2)$ WPIR scheme that satisfies (26) and with MaxL $\rho^{(\text{MaxL})} \leq \rho \leq 1$. Then, the maximum possible WPIR rate is

$$\text{R}_{\text{max}}^{(\text{MaxL})}(\rho) = \left[\frac{5}{2} - 2^\rho\right]^{-1}.$$

Moreover, assuming that only one of the two servers can leak information, then the maximum possible WPIR rate is

$$\text{R}_{\text{max}}^{(\text{MaxL})}(\rho) = \left[2 - 2^{\rho-1}\right]^{-1}. \qquad \Diamond$$

Proof: The achievable scheme for the first assertion is Scheme A with time-sharing. In particular, by applying the time-sharing principle to Example 1, we get the following conditional PMF of $\overline{\boldsymbol{Q}}_1$ given $M = m$,

$P_{\overline{\boldsymbol{Q}}_1 \mid M}(\bar{\boldsymbol{q}}_1 \mid m)$	$m = 1$	$m = 2$
$\bar{\boldsymbol{q}}_1 = (0,0)$	$\dfrac{1-p}{2}$	$\dfrac{1-p}{2}$
$\bar{\boldsymbol{q}}_1 = (0,1)$	$\dfrac{p}{2}$	$\dfrac{1-p}{2}$
$\bar{\boldsymbol{q}}_1 = (1,0)$	$\dfrac{1-p}{2}$	$\dfrac{p}{2}$
$\bar{\boldsymbol{q}}_1 = (1,1)$	$\dfrac{p}{2}$	$\dfrac{p}{2}$

Thus, by definitions we get $2^{\rho^{(\text{MaxL})}} = \frac{1}{2} + (1 - p)$, for $0 \leq p \leq \frac{1}{2}$, and hence it can be seen that $\text{D} = 1 + p = 1 + \frac{3}{2} - 2^{\rho^{(\text{MaxL})}} = \frac{5}{2} - 2^{\rho^{(\text{MaxL})}}$. Note that
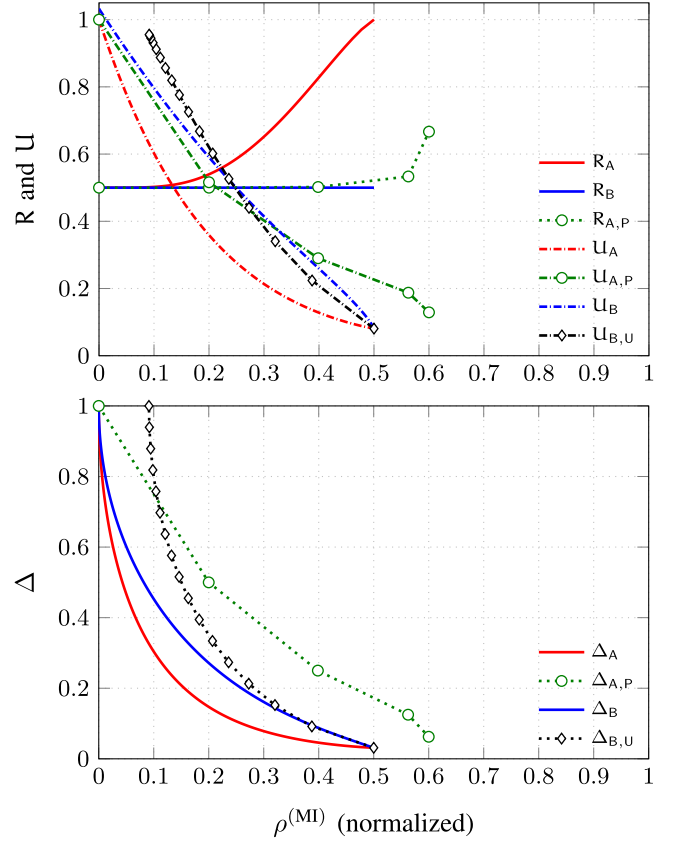

Fig. 2. R, U, and Δ of different WPIR schemes for $M = 32$, as a function of $\rho^{(\text{MI})}$. For $M = 32$, $\text{C}_{M,2}$ is almost equal to $\frac{1}{2}$.

this is also the optimized time-sharing Scheme A under the MaxL metric for $(M, n) = (2, 2)$, where the optimal solution for (20) is $(z_0^*, z_1^*) = (2 - \text{D}, \text{D} - 1)$. The converse part is proven by (30), which gives

$$\frac{\text{D}(\mathscr{C})}{\beta \log_2 |\mathcal{X}|} \geq 1 + \frac{1}{2} - \frac{2^\rho - 1}{1} = \frac{5}{2} - 2^\rho$$

for $(M, n) = (2, 2)$.

Further, the second assertion can be proven by following the same lines as in the poof of Theorem 9. ∎

X. NUMERICAL RESULTS

This section consists of three subsections. Section X-A considers the case of two servers and compares the achievable 4-tuples $(\text{R}, \text{U}, \Delta, \rho^{(\cdot)})$ for the $(M, 2)$ WPIR schemes proposed in Sections IV-A, IV-B.3, V-A, and V-B. Note that as the performance of $\mathscr{C}^{\text{basic}}$ in Section III is inferior to the scheme in Section IV-B.3, the results of $\mathscr{C}^{\text{basic}}$ are not presented. Section X-B also focuses on the case of two servers and compares our $(M, 2)$ WPIR scheme proposed in Section VI-A with the scheme proposed in [60] and [61] under the ϵ-privacy metric. Section X-C presents optimized values for the download rate for the time-sharing Scheme A by numerically solving the convex optimization problem in (20) for both the MI and MaxL privacy metrics and comparisons with the converse bounds from Theorems 8 and 10.
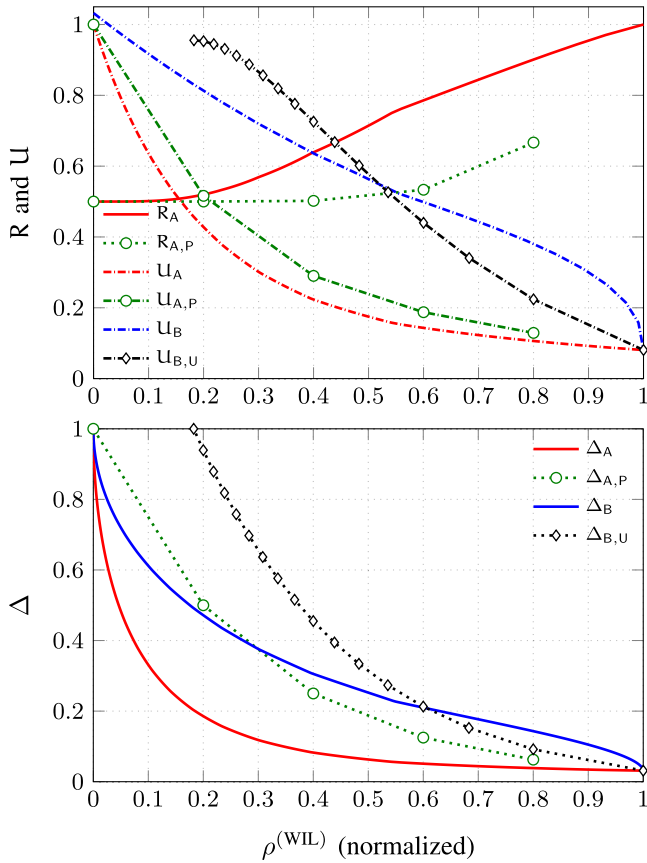
Fig. 3. R, U, and Δ of different WPIR schemes for $M = 32$, as a function of $\rho^{(\text{WIL})}$. Here, $R_B = \frac{1}{2}$ is not plotted.
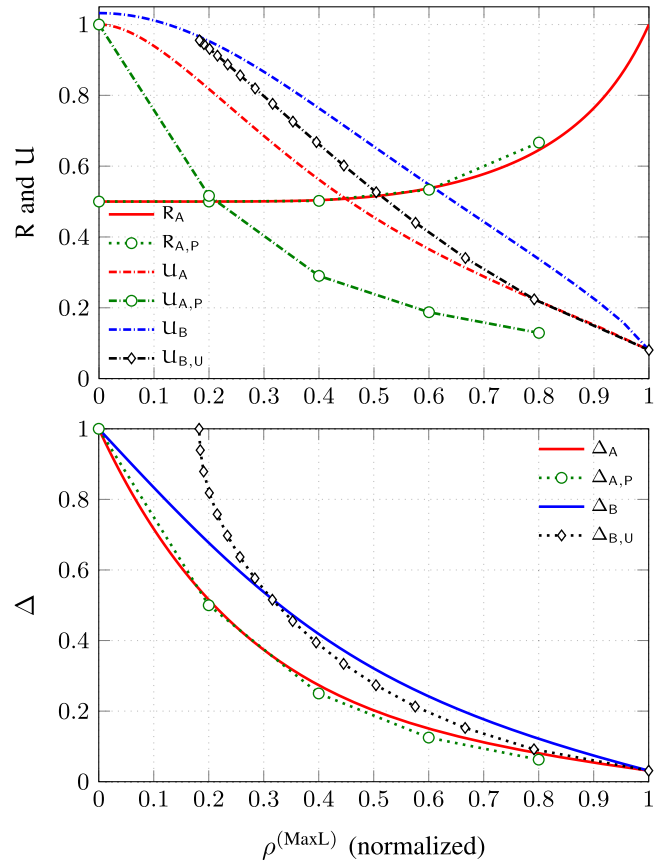

Fig. 4. R, U, and Δ of different WPIR schemes for $M = 32$, as a function of $\rho^{(\text{MaxL})}$. Here, $R_B = \frac{1}{2}$ is not plotted.

A. $(M, 2)$ WPIR Schemes

Four $(M, 2)$ WPIR schemes presented in Sections IV-A, IV-B.3, V-A, and V-B are illustrated. More specifically, the 4-tuples $\left(R_A, U_A, \Delta_A, \rho_A^{(\cdot)} \right)$ achieved by the $(M, 2)$ Scheme A with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to Bernoulli (p), $\left(R_{A,P}, U_{A,P}, \Delta_{A,P}, \rho_{A,P}^{(\cdot)} \right)$ achieved by the (M, n) partition Scheme A using the $\left(\frac{M}{\eta}, n \right)$ Scheme A with $\{S_j\}_{j=1}^{\frac{M}{\eta}-1}$ i.i.d. according to $\mathcal{U}([0:n-1])$ as a subscheme, $\left(\frac{1}{2}, U_B, \Delta_B, \rho_B^{(\cdot)} \right)$ achieved by the $(M, 2)$ Scheme B with $\{S_j\}_{j=1}^{M}$ i.i.d. according to Bernoulli (p), and $\left(\frac{1}{2}, U_{B,U}, \Delta_{B,U}, \rho_{B,U}^{(\cdot)} \right)$ achieved by the $(M, 2)$ Scheme B with $S \sim \mathcal{U}(\mathcal{B}_{w,M})$, are presented for comparison. For the sake of illustration, the information leakage $\rho^{(\text{MI})}$ is normalized by $\log_2 M$ bits, while the upload cost and access complexity are normalized by $2(M-1)$ and M, respectively. $2(M-1)$ and M are the upload cost and access complexity of the PIR capacity-achieving scheme presented in [30] for the case of two servers. The upload cost $2(M-1)$ is optimal among all so-called *decomposable* PIR capacity-achieving schemes [30].[8]

Fig. 2 presents the results of the four WPIR schemes for the case of $M = 32$ files and leakage metric $\rho^{(\text{MI})}$. We can

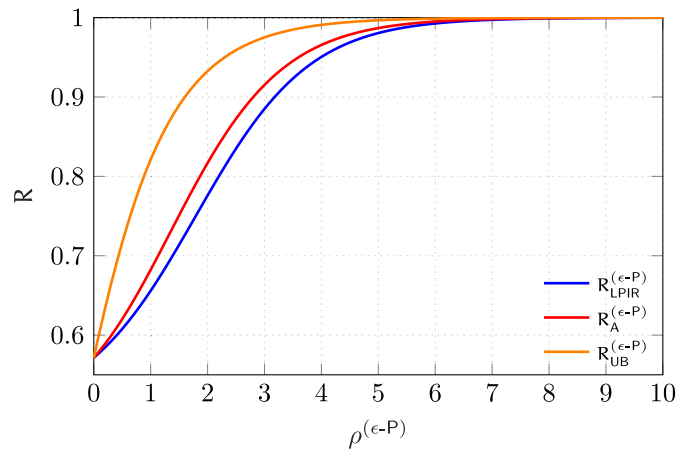[8] Based on [30, Def. 2], all existing PIR schemes in the literature are decomposable.

Fig. 5. The rates $R_{\text{LPIR}}^{(\epsilon\text{-P})}$, $R_A^{(\epsilon\text{-P})}$, and $R_{\text{UB}}^{(\epsilon\text{-P})}$ for $(M, n) = (3, 2)$, as a function of $\rho^{(\epsilon\text{-P})}$.

see that Scheme A yields the best performance in terms of download rate, upload cost, and access complexity for all values of the information leakage. Note that the IR rate of Scheme B with different S is always equal to $\frac{1}{2}$. The results of the four WPIR schemes for WIL and MaxL are provided in Figs. 3 and 4, respectively. For WIL, Scheme A performs best among the four schemes for all values of the information leakage in terms of download rate, upload cost, and access
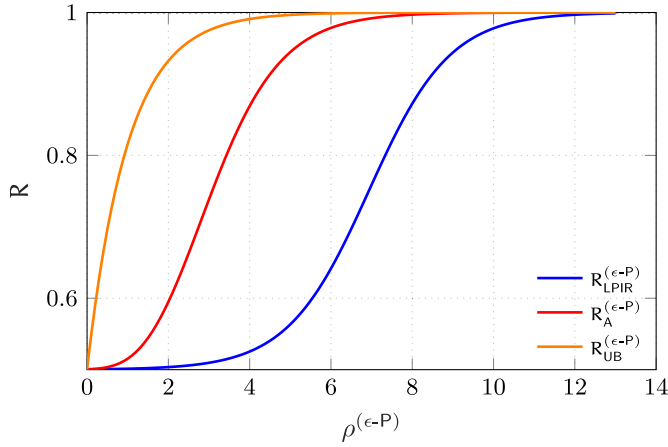
Fig. 6. The rates $R_{LPIR}^{(\epsilon\text{-}P)}$, $R_A^{(\epsilon\text{-}P)}$, and $R_{UB}^{(\epsilon\text{-}P)}$ for $(M, n) = (10, 2)$, as a function of $\rho^{(\epsilon\text{-}P)}$.

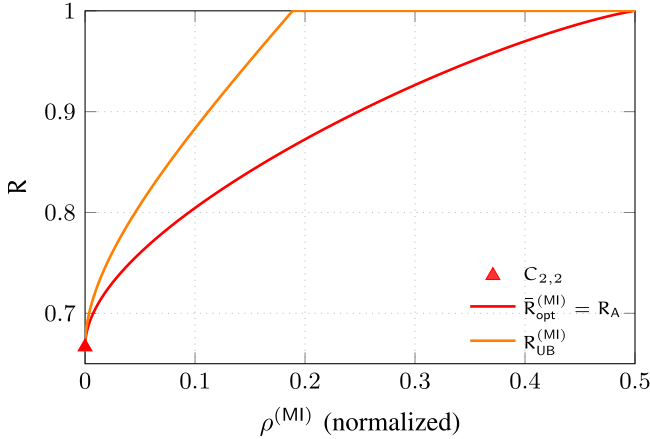Fig. 8. The optimized rate $\bar{R}_{opt}^{(MI)}$, R_A, and $R_{A,P}$ for $(M, n) = (6, 2)$, as a function of $\rho^{(MI)}$.

Fig. 7. The optimized rate $\bar{R}_{opt}^{(MI)}$ for the time-sharing Scheme A and $R_{UB}^{(MI)}$ for $(M, n) = (2, 2)$, as a function of $\rho^{(MI)}$.

Fig. 9. The optimized rate $\bar{R}_{opt}^{(MI)}$ and $R_{A,P}$ for $(M, n) = (6, 3)$, as a function of $\rho^{(MI)}$.

complexity. However, for MaxL the partition Scheme A (from Theorem 4) has a comparable performance to Scheme A with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to Bernoulli (p) (from Theorem 3) for both download rate and access complexity. In particular, for $\rho^{(MaxL)} = 0.8$ it exhibits a slightly higher download rate, whereas for $0.2 \leq \rho^{(MaxL)} \leq 0.8$ it achieves a lower access complexity. On the other hand, it yields a significantly lower upload cost for all values of the information leakage.

B. ϵ-Privacy for $(M, 2)$ WPIR Schemes

In this subsection, the achievable rates and the upper bound under the ϵ-privacy metric presented in Section VI are demonstrated. We evaluate the rate $R_{LPIR}^{(\epsilon\text{-}P)}$, the upper bound $R_{UB}^{(\epsilon\text{-}P)}$, and the rate $R_A^{(\epsilon\text{-}P)}$ in (15), (16), and (17), respectively. Fig. 5 presents the results for the case of $M = 3$ files. In Fig. 6, the corresponding curves for $(M, n) = (10, 2)$ are provided. It can be seen that our proposed $(M, 2)$ Scheme A outperforms the leaky PIR scheme presented in [60], [61], in terms of download rate. Note that for the special case of $M = 2$ files, we have $R_{LPIR}^{(\epsilon\text{-}P)} = R_A^{(\epsilon\text{-}P)}$. Moreover, the converse bound $R_{UB}^{(\epsilon\text{-}P)}$ is in general not tight.

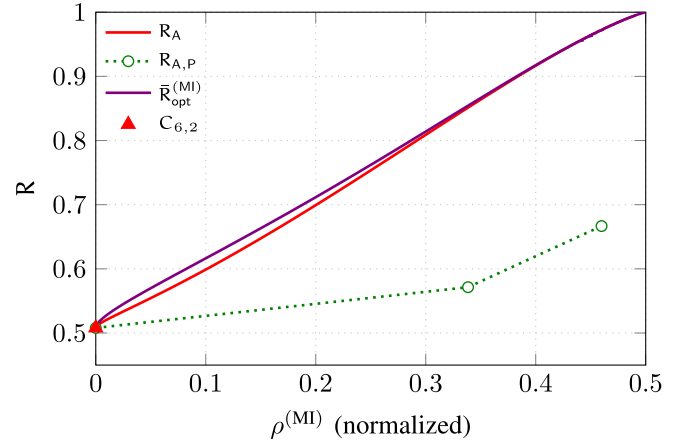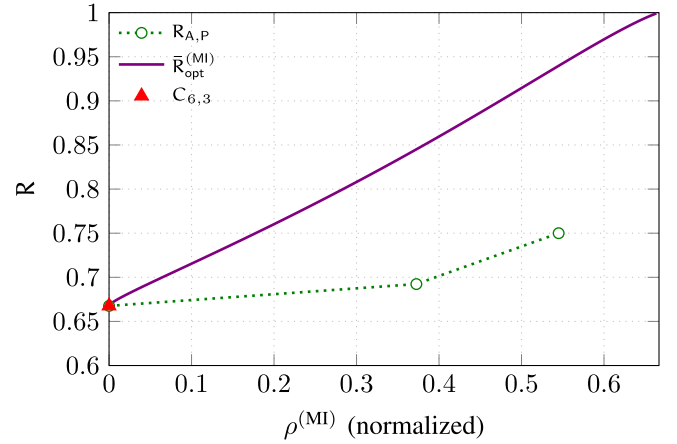### C. Optimized Rates for the Time-Sharing Scheme A

In this subsection, we give the maximum download rate under a leakage constraint for the time-sharing Scheme A described in Section VII with both the MI and MaxL privacy metrics. Since for both metrics optimizing the download rate is a convex problem (see (20)), the optimal solutions can be obtained by using the CVXPY Python-embedded modeling language for convex optimization problems [76], [77]. The optimal corresponding rate obtained from (20) is denoted by $\bar{R}_{opt}^{(\cdot)}$. Unless specified otherwise, all solutions are numerically computed. Moreover, if the converse bound $R_{UB}^{(\cdot)}$ (see Theorems 8 and 10) is trivial for all leakage constraints, i.e., $R_{UB}^{(\cdot)} \geq 1$, we do not include it in the figures.

Under the MI privacy metric, Fig. 7 compares the optimal rate-leakage tradeoff curve for the canonical case of $(M, n) = (2, 2)$ to the converse bound $R_{UB}^{(MI)}$ from Theorem 8, which shows that in general it is not tight. We remark that the optimal curve is equal to the curve presented in Example 1, and it can also be shown that the analytical optimal solution can be derived directly from (20). In Fig. 8, for $(M, n) = (6, 2)$, the download rate from Theorems 3 and 4 is plotted as a function
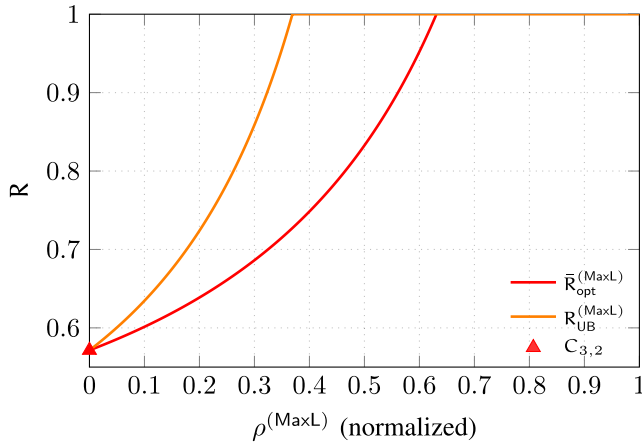
Fig. 10. The optimized rate $\bar{R}_{opt}^{(MaxL)}$ for the time-sharing Scheme A and $R_{UB}^{(MaxL)}$ for $(M, n) = (3, 2)$, as a function of $\rho^{(MaxL)}$.
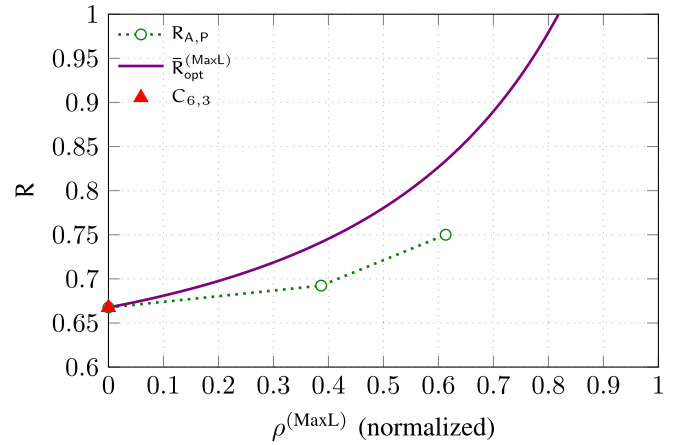

Fig. 12. The optimized rate $\bar{R}_{opt}^{(MaxL)}$ and $R_{A,P}$ for $(M, n) = (6, 3)$, as a function of $\rho^{(MaxL)}$.

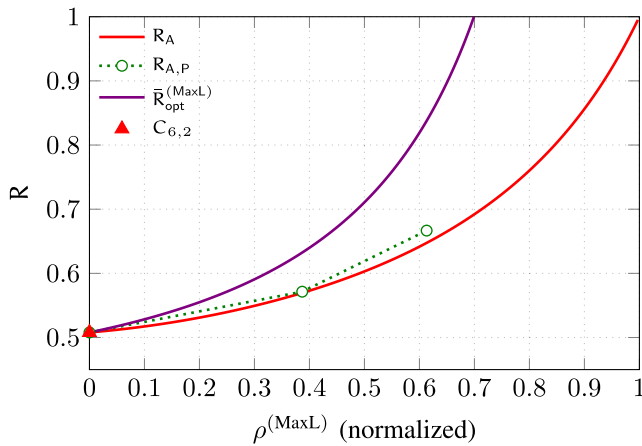
Fig. 11. The optimized rate $\bar{R}_{opt}^{(MaxL)}$, R_A, and $R_{A,P}$ for $(M, n) = (6, 2)$, as a function of $\rho^{(MaxL)}$.

of the information leakage, together with the optimal download rate $\bar{R}_{opt}^{(MI)}$ for the time-sharing Scheme A obtained from (20). The comparisons show that Scheme A with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to Bernoulli (p) (from Theorem 3) exhibits a download rate that is close to being optimal. On the other hand, partition Scheme A (from Theorem 4) performs quite far from the optimal tradeoff curve. In Fig. 9, the corresponding curves for $(M, n) = (6, 3)$ (excluding the curve from Theorem 3, which assumes $n = 2$) are presented. Again, partition Scheme A performs far from the optimal tradeoff curve. Note that for both $(M, n) = (6, 2)$ and $(6, 3)$ the converse bound from Theorem 8 is trivial.

In Figs. 10 to 12, the corresponding curves for the MaxL privacy metric are depicted. In particular, the figures show results for $(M, n) = (3, 2)$, $(6, 2)$, and $(6, 3)$, respectively. Note that with the MaxL privacy metric the converse bound from Theorem 10 is tight for the canonical case of $(M, n) = (2, 2)$ (see Theorem 11). Hence, in contrast to Fig. 7, where $(M, n) = (2, 2)$, we use $(M, n) = (3, 2)$ in Fig. 10. From Fig. 11, for $(M, n) = (6, 2)$, Scheme A with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to Bernoulli (p) (from Theorem 3) exhibits a lower

download rate than partition Scheme A (from Theorem 4), which is in contrast to the case of MI leakage where partition Scheme A performs significantly worse (see Fig. 8). Moreover, the gap to the optimized rate $\bar{R}_{opt}^{(MaxL)}$ is higher than with the MI privacy metric. For $(M, n) = (6, 3)$, Fig. 12 shows that the gap in download rate between the optimized rate $\bar{R}_{opt}^{(MaxL)}$ and the rate from partition Scheme A is smaller than with the MI privacy metric, which indicates that partition Scheme A performs better with the MaxL privacy metric than with the MI privacy metric. For both $(M, n) = (6, 2)$ and $(6, 3)$ (as for the MI privacy metric) the converse bound from Theorem 10 is trivial.

XI. CONCLUSION

We presented the first study of the tradeoffs that can be achieved by relaxing the perfect privacy requirement of PIR, referred to as WPIR, for the case of multiple replicated non-colluding servers. Two WPIR schemes based on two different PIR protocols, named Scheme A and Scheme B, and a family of schemes based on partitioning were proposed. The proposed model shows that by relaxing the perfect privacy requirement, the download rate, the upload cost, and the access complexity can be improved. In addition, we showed that Scheme A achieves an improved download rate compared to the leaky PIR scheme proposed by Samy *et al.* under the ϵ-privacy metric. Under the MI and MaxL privacy metrics and with a practical restriction on the alphabet size of queries and answers, we provided an information-theoretic converse bound on the download rate. For the MaxL privacy metric and for two servers and two files, the converse bound is tight, giving the WPIR capacity in this special case. Numerous numerical results were presented, comparing the performance of the proposed schemes and their gap to the new converse bound.

Many interesting directions can be studied as future work. First of all, the derivation of a better converse bound on the download rate, as well as on the upload cost and the access complexity, is worth further investigation for general cases of (M, n). On the other hand, practical variants of WPIR include colluding, Byzantine, and unresponsive servers, are

all important topics for future research, as well as WPIR for coded DSSs and WPIR with secure storage.

APPENDIX A
PROOF OF THEOREM 2

Without loss of generality, denote the requested file index M by $M \equiv (M_J, J)$, where M_J denotes the requested file index in the J-th partition. The MI based leakage at the l-th server, $l \in [n]$, is given as

$$
\begin{aligned}
\mathsf{I}(M; \boldsymbol{Q}_l) &= \mathsf{H}(M) - \mathsf{H}(M|\boldsymbol{Q}_l) \\
&= \mathsf{H}(M_J, J) - \mathsf{H}(M_J, J|\widetilde{\boldsymbol{Q}}_l, J) \\
&= \mathsf{H}(J) + \mathsf{H}(M_J|J) - \mathsf{H}(M_J|\widetilde{\boldsymbol{Q}}_l, J) \\
&\overset{(a)}{=} \mathsf{H}(J) + \mathsf{H}(M_J) - \mathsf{H}(M_J|\widetilde{\boldsymbol{Q}}_l) \\
&= \log_2 \eta + \mathsf{I}(M_J; \widetilde{\boldsymbol{Q}}_l),
\end{aligned}
\tag{31}
$$

where (a) follows since M_J and J are assumed to be uniform RVs, and hence knowing J does not reveal any information about the requested file index M_J. Using (31) in (4) gives $\rho^{(\mathsf{MI})}(\mathscr{C})$. Using a similar argument as above, the expressions for $\mathsf{R}(\mathscr{C})$, $\mathsf{U}(\mathscr{C})$, $\Delta(\mathscr{C})$, $\rho^{(\mathsf{WIL})}(\mathscr{C})$, and $\rho^{(\mathsf{MaxL})}(\mathscr{C})$ can be derived accordingly.

APPENDIX B
PROOF OF LEMMA 2

Observe that if $\{S_j\}_{j=1}^{\mathsf{M}-1}$ are i.i.d. according to $\mathcal{U}([0:n-1])$, then for every $m, m' \in [\mathsf{M}]$ with $m \neq m'$, it holds that

$$
\Pr[\boldsymbol{Q}_l = \boldsymbol{q}_l | M = m] = \Pr[\boldsymbol{Q}_l = \boldsymbol{q}_l | M = m'] = \left(\frac{1}{n}\right)^{\mathsf{M}-1}
$$

for all $\boldsymbol{q}_l \in \mathcal{Q}_l$, $l \in [n]$. Moreover, from the answer construction of (11), the IR rate is

$$
\begin{aligned}
\mathsf{R} &= \frac{\beta \log_2 2}{\log_2 2 \sum_{l=1}^{n} \mathbb{E}_{\boldsymbol{Q}_l}[\ell_l(\boldsymbol{Q}_l)]} \\
&= \frac{n-1}{[1 - P_{\boldsymbol{Q}_1}(\boldsymbol{0})] + \sum_{l=2}^{n} \mathbb{E}_{\boldsymbol{Q}_l}[\ell_l(\boldsymbol{Q}_l)]} \\
&= \frac{n-1}{\left(1 - \frac{1}{n^{\mathsf{M}-1}}\right) + (n-1)} = \frac{1 - \frac{1}{n}}{1 - \frac{1}{n^{\mathsf{M}}}},
\end{aligned}
$$

which is equal to the n-server PIR capacity for M files.

APPENDIX C
PROOF OF THEOREM 3

From the theorem statement, the entries $\{S_j\}_{j=1}^{\mathsf{M}-1}$ of the random strategy $\boldsymbol{S}$ are assumed to be i.i.d. according to Bernoulli (p), $0 \leq p \leq \frac{1}{2}$. Hence, $P_{\boldsymbol{Q}_1}(\boldsymbol{0}) = P_{\boldsymbol{S}}(\boldsymbol{0}) = (1-p)^{\mathsf{M}-1}$, and we have

$$
\mathsf{R}(\mathscr{C}_\mathsf{A}) = \frac{1}{[1 - (1-p)^{\mathsf{M}-1}] + 1}
$$

from the general formula in (13).

For the upload cost, access complexity, and the information leakage metrics, we first derive the PMF of $\boldsymbol{Q}_l$, $l = 1, 2$. Let

us consider a query $\boldsymbol{q}_l$ to the l-th server that has $w_\mathsf{H}(\boldsymbol{q}_l) = w$. Due to the query generation, we have

$$
P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m) =
\begin{cases}
(1-p)^{\mathsf{M}-w-1}p^w & \text{if } m \in [\mathsf{M}] \setminus \chi(\boldsymbol{q}_l), \\
(1-p)^{\mathsf{M}-w}p^{w-1} & \text{if } m \in \chi(\boldsymbol{q}_l).
\end{cases}
\tag{32}
$$

By using the law of total probability, we obtain

$$
\begin{aligned}
P_{\boldsymbol{Q}_l}(\boldsymbol{q}_l) &= \sum_{m'=1}^{\mathsf{M}} \frac{1}{\mathsf{M}} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m') \\
&= \frac{1}{\mathsf{M}} \left[\binom{\mathsf{M}-w}{1} \cdot (1-p)^{\mathsf{M}-w-1} \cdot p^w \right. \\
&\qquad\left. + \binom{w}{1} \cdot (1-p)^{\mathsf{M}-w} \cdot p^{w-1} \right] \\
&= f(w, p).
\end{aligned}
$$

From the query generation (see Section IV-.1), it follows that $w_\mathsf{H}(\boldsymbol{q}_1)$ must be even and $w_\mathsf{H}(\boldsymbol{q}_2)$ must be odd for the case of $n = 2$ servers, hence, the upload cost is equal to

$$
\begin{aligned}
\mathsf{U}(\mathscr{C}_\mathsf{A}) &= \mathsf{H}(\boldsymbol{Q}_1) + \mathsf{H}(\boldsymbol{Q}_2) \\
&= -\sum_{w=0}^{\mathsf{M}} \binom{\mathsf{M}}{w} f(w, p) \log_2\big(f(w, p)\big).
\end{aligned}
$$

Further, by the definition in (6), the access complexity $\Delta(\mathscr{C}_\mathsf{A}) = \Delta_\mathsf{A}$ follows.

Moreover, we have

$$
\mathsf{H}(\boldsymbol{Q}_l|M) = \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \mathsf{H}(\boldsymbol{S}) = (\mathsf{M}-1)\mathsf{H}_\mathsf{b}(p),
\tag{33}
$$

where (33) holds by the query generation, the fact that the entropy of i.i.d. RVs is equal to the sum of the individual entropies, and $S_j \sim$ Bernoulli (p). Hence, we obtain

$$
\begin{aligned}
\rho^{(\mathsf{MI})} &= \frac{\mathsf{H}(\boldsymbol{Q}_1) - \mathsf{H}(\boldsymbol{Q}_1|M) + \mathsf{H}(\boldsymbol{Q}_2) - \mathsf{H}(\boldsymbol{Q}_2|M)}{2} \\
&= \frac{1}{2}\mathsf{U}_\mathsf{A} - (\mathsf{M}-1)\mathsf{H}_\mathsf{b}(p).
\end{aligned}
$$

For the WIL metric, applying Bayes' rule, given $M = m$ and a query $\boldsymbol{q}_l$ with $w_\mathsf{H}(\boldsymbol{q}_l) = w$, we get

$$
P_{M|\boldsymbol{Q}_l}(m|\boldsymbol{q}_l) = \frac{P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)}{\sum_{m'=1}^{\mathsf{M}} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m')},
\tag{34}
$$

where (34) holds since the requested file index M is assumed to be uniformly distributed. Finally, since the requested index $m \in [\mathsf{M}]$ either belongs to $[\mathsf{M}] \setminus \chi(\boldsymbol{q}_l)$ or $\chi(\boldsymbol{q}_l)$, it is not too difficult to see that

$$
\begin{aligned}
&P_{M|\boldsymbol{Q}_l}(m|\boldsymbol{q}_l) \\
&= \begin{cases}
\dfrac{(1-p)^{\mathsf{M}-w-1}p^w}{(\mathsf{M}-w)(1-p)^{\mathsf{M}-w-1}p^w + w(1-p)^{\mathsf{M}-w}p^{w-1}} \\
\qquad\qquad\qquad\qquad \text{if } m \in [\mathsf{M}] \setminus \chi(\boldsymbol{q}_l), \\
\dfrac{(1-p)^{\mathsf{M}-w}p^{w-1}}{(\mathsf{M}-w)(1-p)^{\mathsf{M}-w-1}p^w + w(1-p)^{\mathsf{M}-w}p^{w-1}} \\
\qquad\qquad\qquad\qquad \text{if } m \in \chi(\boldsymbol{q}_l),
\end{cases}
\end{aligned}
\tag{35}
$$

where (35) follows from (32). Note that to compute the entropy $H(M|\boldsymbol{Q}_l = \boldsymbol{q}_l)$, we only need to know the conditional PMF of M given $\boldsymbol{Q}_l = \boldsymbol{q}_l$, hence, we can introduce a new RV $M_w \equiv M_{\boldsymbol{q}_l}$ with $w_H(\boldsymbol{q}_l) = w$ that has an equivalent PMF defined as (14). This then gives $\rho_A^{(\text{WIL})}$.

Finally, we derive the expression for $\rho_A^{(\text{MaxL})}$. From (32), it follows that

$$\max_{m \in [M]} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)$$
$$= \begin{cases} (1-p)^{M-1} & \text{if } w_H(\boldsymbol{q}_l) = 0, \\ (1-p)^{M-w}p^{w-1} & \text{otherwise.} \end{cases}$$

Moreover,

$$p\left(2^{\text{MaxL}(M;\boldsymbol{Q}_1)} - 2^{\text{MaxL}(M;\boldsymbol{Q}_2)}\right)$$
$$= p(1-p)^{M-1} + \sum_{\substack{w \in [2:M] \\ w:\text{ even}}} \binom{M}{w}(1-p)^{M-w}p^{w}$$
$$\quad - \sum_{\substack{w \in [M] \\ w:\text{ odd}}} \binom{M}{w}(1-p)^{M-w}p^{w}$$
$$= \sum_{w \in [0:M]} \binom{M}{w}(1-p)^{M-w}(-p)^{w}$$
$$\quad - (1-p)^{M} + p(1-p)^{M-1}$$
$$= \left[(1-p)-p\right]^{M} - (1-p)^{M-1}(1-2p) < 0, \quad (36)$$

where (36) follows by binomial expansion. Since (36) is nonpositive when $0 \le p \le \frac{1}{2}$, the expression for $\rho_A^{(\text{MaxL})}$ follows immediately.

APPENDIX D
PROOF OF THEOREM 4

Since the $\left(\frac{M}{\eta}, n\right)$ Scheme A with $\{S_j\}_{j=1}^{\frac{M}{\eta}-1}$ i.i.d. according to $\mathcal{U}([0:n-1])$ is used as a subscheme, from Section IV-B.1, we have

$$P_{\boldsymbol{Q}_l|M}\left(\boldsymbol{q}_l|(j-1)\frac{M}{\eta}+m'\right) = \left(n^{\frac{M}{\eta}-1}\right)^{-1}, \quad (37)$$

$j \in [\eta]$, $m' \in \left[\frac{M}{\eta}\right]$, $\forall \boldsymbol{q}_l \in \mathcal{Q}_l$, $l \in [n]$. Similar to Appendix C, we obtain

$$P_{\boldsymbol{Q}_l}(\boldsymbol{q}_l) \overset{(a)}{=} \begin{cases} \frac{1}{M}\sum_{m=1}^{M} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m) & \text{if } \boldsymbol{q}_l = \boldsymbol{0}, \\ \frac{1}{M}\sum_{m \in \mathcal{P}_j} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m) & \text{otherwise} \end{cases}$$
$$= \begin{cases} \frac{M}{M}\left(n^{\frac{M}{\eta}-1}\right)^{-1} & \text{if } \boldsymbol{q}_l = \boldsymbol{0}, \\ \frac{\frac{M}{\eta}}{M}\left(n^{\frac{M}{\eta}-1}\right)^{-1} & \text{otherwise} \end{cases}$$
$$= \begin{cases} \left(n^{\frac{M}{\eta}-1}\right)^{-1} & \text{if } l = 1 \text{ and } \boldsymbol{q}_l = \boldsymbol{0}, \\ \left(\eta \cdot n^{\frac{M}{\eta}-1}\right)^{-1} & \text{otherwise,} \end{cases}$$

where (a) holds since in Scheme A the user can send the all-zero query to the first server to request any file in any partition group. Since $P_{\boldsymbol{Q}_1}(\boldsymbol{0}) = \left(n^{\frac{M}{\eta}-1}\right)^{-1}$, using (13) gives

$$R(\mathscr{C}_A^{\text{part}}) = \frac{n-1}{\left[1 - \left(n^{\frac{M}{\eta}-1}\right)^{-1}\right] + (n-1)} = R_{\text{A,P}}.$$

For the upload cost, since there are $\eta\left[n^{\frac{M}{\eta}-1} - 1\right]$ equally-likely nonzero queries in $\mathcal{Q}_1$, it can be shown that

$$H(\boldsymbol{Q}_1) = \frac{1}{n^{\frac{M}{\eta}-1}} \log_2\left(n^{\frac{M}{\eta}-1}\right)$$
$$\quad + \eta\left[n^{\frac{M}{\eta}-1} - 1\right] \cdot \frac{1}{\eta \cdot n^{\frac{M}{\eta}-1}} \log_2\left(\eta \cdot n^{\frac{M}{\eta}-1}\right)$$
$$= \left(\frac{M}{\eta} - 1\right)\log_2 n + \log_2 \eta - \frac{\log_2 \eta}{n^{\frac{M}{\eta}-1}}.$$

Similarly, we have $H(\boldsymbol{Q}_l) = \left(\frac{M}{\eta} - 1\right)\log_2 n + \log_2 \eta$ for all $l \in [2:n]$. This then gives the expression for $U(\mathscr{C}_A^{\text{part}}) = U_{\text{A,P}}$.

For the access complexity, recall that Scheme A partitions all M files into equally-sized $\frac{M}{\eta}$ groups and Scheme A with $\{S_j\}_{j=1}^{\frac{M}{\eta}-1}$ i.i.d. according to $\mathcal{U}([0:n-1])$ is treated as a subscheme for each partition group. Thus, we have

$$\Delta(\mathscr{C}_A^{\text{part}}) = \sum_{l=1}^{n} \mathbb{E}_{\boldsymbol{Q}_l}\left[\delta_l(\boldsymbol{Q}_l)\right]$$
$$= \sum_{l=1}^{n} \sum_{\boldsymbol{q}_l \in \mathcal{Q}_l} w_H(\boldsymbol{q}_l) P_{\boldsymbol{Q}_l}(\boldsymbol{q}_l)$$
$$\overset{(a)}{=} \sum_{\substack{w \in \left[0:\frac{M}{\eta}\right] \\ w>0}} w \cdot \eta\binom{\frac{M}{\eta}}{w}(n-1)^{w}\frac{1}{\eta \cdot n^{\frac{M}{\eta}-1}}$$
$$= \sum_{\substack{w \in \left[0:\frac{M}{\eta}\right] \\ w>0}} w\binom{\frac{M}{\eta}}{w}(n-1)^{w-1}\frac{n-1}{n^{\frac{M}{\eta}-1}}$$
$$\overset{(b)}{=} \frac{n-1}{n^{\frac{M}{\eta}-1}} \cdot \left(\frac{M}{\eta} \cdot n^{\frac{M}{\eta}-1}\right) = (n-1)\frac{M}{\eta},$$

where (a) follows since for each partition group, (10) indicates that an $\left(\frac{M}{\eta}, n\right)$ Scheme A consists of in total $\binom{\frac{M}{\eta}}{w}(n-1)^{w}$ nonzero queries with Hamming weight w in $\bigcup_{l=1}^{n} \mathcal{Q}_l$; (b) is due to the fact that $\sum_{h=0}^{z} h\binom{z}{h} \cdot x^{h-1} = \frac{d}{dx}(1+x)^{z} = \frac{d}{dx}\left(\sum_{h=0}^{z}\binom{z}{h}x^{h}\right) = z(1+x)^{z-1}$ for some $z \in \mathbb{N}$.

For the information leakage metric $\rho^{(\text{MI})}$, similar to (33), we have

$$H(\boldsymbol{Q}_l|M) = \left(\frac{M}{\eta} - 1\right)H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \left(\frac{M}{\eta} - 1\right)\log_2 n,$$

$l \in [n]$, and hence $\rho^{(\text{MI})}(\mathscr{C}_A^{\text{part}}) = \frac{U_{\text{A,P}}}{n} - \left(\frac{M}{\eta} - 1\right)\log_2 n = \rho_{\text{A,P}}^{(\text{MI})}$ is achievable.

Under the WIL metric $\rho^{(\text{WIL})}$, if $w_H(\boldsymbol{q}_l) = 0$, we obtain $P_{M|\boldsymbol{Q}_l}(m|\boldsymbol{q}_l) = \frac{1}{M}$, $\forall m \in [M]$, while $P_{M|\boldsymbol{Q}_l}\left((j-1)\frac{M}{\eta} + m'|\boldsymbol{q}_l\right) = \frac{1}{\frac{M}{\eta}}$ for $w_H(\boldsymbol{q}_l) > 0$, $j \in [\eta]$, $m' \in \left[\frac{M}{\eta}\right]$. Therefore, we obtain

$$H(M|\boldsymbol{Q}_l = \boldsymbol{q}_l) = \begin{cases} \log_2 M & \text{if } \boldsymbol{q}_l = \boldsymbol{0}, \\ \log_2\left(\frac{M}{\eta}\right) & \text{otherwise.} \end{cases}$$

Because $\log_2\left(\frac{M}{\eta}\right) \le \log_2 M$ for $\eta \in [M-1]$, the achievability of $\rho^{(\text{WIL})}(\mathscr{C}_A^{\text{part}}) = \rho_{A,P}^{(\text{WIL})}$ follows.

For the privacy metric $\rho^{(\text{MaxL})}$, it follows from (37) that

$$2^{\text{MaxL}(M;\boldsymbol{Q}_l)}$$

$$= \sum_{\boldsymbol{q}_l \in \mathcal{Q}_l} \max_{m\in[M]} P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)$$

$$= |\mathcal{Q}_l| \cdot \left(n^{\frac{M}{\eta}-1}\right)^{-1}$$

$$= \begin{cases} \left(1 + \eta\left(n^{\frac{M}{\eta}-1}-1\right)\right) \cdot \left(n^{\frac{M}{\eta}-1}\right)^{-1} < \eta & \text{for } l = 1, \\ \eta \cdot n^{\frac{M}{\eta}-1} \cdot \left(n^{\frac{M}{\eta}-1}\right)^{-1} = \eta & \text{for } l \in [2:n], \end{cases}$$

which implies that $\rho^{(\text{MaxL})}(\mathscr{C}_A^{\text{part}}) = \log_2 \eta = \rho_{A,P}^{(\text{MaxL})}$.

APPENDIX E
PROOF OF THEOREM 6

When the user requests the M-th file, the $(M,2)$ Scheme B sends $\boldsymbol{Q}_1 = \boldsymbol{S} + \boldsymbol{v}_M$ and $\boldsymbol{Q}_2 = \boldsymbol{S}$ to the respective servers, where $\boldsymbol{v}_M$ is the M-th M-dimensional unit vector. As the random strategy $\boldsymbol{S} \sim \mathcal{U}(\mathcal{B}_{w,M})$ is taken, it is not too hard to see that

$$P_{\boldsymbol{Q}_1}(\boldsymbol{q}_1) = \begin{cases} \frac{w+1}{\binom{M}{w}M} & \text{if } \boldsymbol{q}_1 \in \mathcal{B}_{w+1,M}, \\ \frac{M-(w-1)}{\binom{M}{w}M} & \text{if } \boldsymbol{q}_1 \in \mathcal{B}_{w-1,M}, \end{cases}$$

$\Pr[\boldsymbol{Q}_2 = \boldsymbol{q}_2] = \frac{1}{\binom{M}{w}}$, $\boldsymbol{q}_2 \in \mathcal{B}_{w,M}$, and $M \perp\!\!\!\perp \boldsymbol{Q}_2$.

Since $\mathsf{H}(\boldsymbol{Q}_l|M) = \mathsf{H}(\boldsymbol{S}) = \log_2\binom{M}{w}$, the results of $\mathsf{U}(\mathscr{C}_B)$, $\Delta(\mathscr{C}_B)$, and $\rho^{(\text{MI})}(\mathscr{C}_B)$ with $\boldsymbol{S} \sim \mathcal{U}(\mathcal{B}_{w,M})$ can be determined by a simple deduction.

Moreover, one can also show that

$$\mathsf{H}(M|\boldsymbol{Q}_1 = \boldsymbol{q}_1) = \begin{cases} \log_2(w+1) & \text{if } \boldsymbol{q}_1 \in \mathcal{B}_{w+1,M}, \\ \log_2(M-(w-1)) & \text{if } \boldsymbol{q}_1 \in \mathcal{B}_{w-1,M}. \end{cases}$$

On the other hand, it can be seen that for $\boldsymbol{q}_1 \in \mathcal{B}_{w+1,M}$,

$$P_{\boldsymbol{Q}_1|M}(\boldsymbol{q}_1|m) = \begin{cases} \frac{1}{\binom{M}{w}} & \text{if } m \in \chi(\boldsymbol{q}_1), \\ 0 & \text{otherwise}, \end{cases}$$

and for $\boldsymbol{q}_1 \in \mathcal{B}_{w-1,M}$,

$$P_{\boldsymbol{Q}_1|M}(\boldsymbol{q}_1|m) = \begin{cases} \frac{1}{\binom{M}{w}} & \text{if } m \in [M] \setminus \chi(\boldsymbol{q}_1), \\ 0 & \text{otherwise}. \end{cases}$$

Therefore, from the above we obtain the expressions for $\rho_{B,U}^{(\text{WIL})}$ and $\rho_{B,U}^{(\text{MaxL})}$.

APPENDIX F
PROOF OF THEOREM 7

From (32) in Appendix C, we know that

$$\max_{m,m'\in[M]} \frac{P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m)}{P_{\boldsymbol{Q}_l|M}(\boldsymbol{q}_l|m')}$$

$$= \begin{cases} \max\left\{\frac{(1-p)^{M-w-1}p^w}{(1-p)^{M-w}p^{w-1}}, \frac{(1-p)^{M-w}p^{w-1}}{(1-p)^{M-w-1}p^w}\right\} \\ \qquad\qquad\qquad \text{if } w_H(\boldsymbol{q}_l) = w \in [M-1], \\ 1 \qquad\qquad\quad\ \text{if } w_H(\boldsymbol{q}_l) = 0 \text{ or } M, \end{cases}$$

for a given query $\boldsymbol{q}_l$, for an (M,n) Scheme A $\mathscr{C}_A$ with $\{S_j\}_{j=1}^{M-1}$ i.i.d. according to Bernoulli (p), $0 \le p \le \frac{1}{2}$. Thus, it can be easily seen that the ϵ-privacy leakage is equal to

$$\rho^{(\epsilon\text{-P})}(\mathscr{C}_A) = \ln\left(\max\left\{\frac{p}{1-p}, \frac{1-p}{p}, 1\right\}\right)$$

$$= \ln\left(\frac{1-p}{p}\right).$$

Hence, to satisfy the leakage constraint $\rho^{(\epsilon\text{-P})} \le \rho$, we require

$$\frac{1-p}{p} \le e^\rho,$$

which gives the inequality $(1+e^\rho)^{-1} \le p \le \frac{1}{2}$ for any $\rho \ge 0$. Finally, to complete the proof, we simply pick $p_\rho \triangleq (1+e^\rho)^{-1}$, and substitute it into (13).

APPENDIX G
PROOF OF LEMMA 6

The lemma can be shown by combining Lemma 5 with a similar approach to the one of the converse proofs given in the information theory literature for PIR, see, e.g., [7], [60], [78]. To make the paper self-contained, we repeat some basic steps here.

The first objective is to find an upper bound on the absolute value of the entropy difference $\mathsf{H}(A_l^{(m)} \mid \boldsymbol{Q}_l^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}) - \mathsf{H}(A_l^{(m')} \mid \boldsymbol{Q}_l^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})$ subject to $\mathsf{I}(M;\boldsymbol{Q}_l) \le \rho_l$, $\forall l \in [n]$, where $m, m' \notin \mathcal{M} \subsetneq [M-1]$.

Observe that

$$\left|\mathsf{H}(A_l^{(m)} \mid \boldsymbol{Q}_l^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}) \right.$$
$$\left. - \mathsf{H}(A_l^{(m')} \mid \boldsymbol{Q}_l^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})\right|$$
$$= \left|\mathsf{H}(A_l^{(m)}, \boldsymbol{Q}_l^{(m)} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}) - \mathsf{H}(\boldsymbol{Q}_l^{(m)} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})\right.$$
$$- \mathsf{H}(A_l^{(m')}, \boldsymbol{Q}_l^{(m')} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})$$
$$\left. + \mathsf{H}(\boldsymbol{Q}_l^{(m')} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})\right|$$
$$\le \left|\mathsf{H}(A_l^{(m)}, \boldsymbol{Q}_l^{(m)} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})\right.$$
$$\left. - \mathsf{H}(A_l^{(m')}, \boldsymbol{Q}_l^{(m')} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})\right|$$
$$+ \left|\mathsf{H}(\boldsymbol{Q}_l^{(m)} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}) - \mathsf{H}(\boldsymbol{Q}_l^{(m')} \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)})\right|$$
$$= \left|\mathsf{H}(A_l, \boldsymbol{Q}_l \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, M = m)\right.$$
$$\left. - \mathsf{H}(A_l, \boldsymbol{Q}_l \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, M = m')\right|$$
$$+ \left|\mathsf{H}(\boldsymbol{Q}_l \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, M = m)\right.$$
$$\left. - \mathsf{H}(\boldsymbol{Q}_l \mid \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, M = m')\right| \tag{38}$$
$$\le \epsilon^{\text{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\text{MI}}(\mathcal{Q}_l), \tag{39}$$

where the inequality (39) can be justified as follows. Using (2) and the Markov chain $M \multimap \boldsymbol{Q}_l \multimap \boldsymbol{A}_l$, we have

$$
\begin{aligned}
& \mathsf{I}\big(M; \boldsymbol{Q}_l, \boldsymbol{A}_l \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& = \mathsf{I}\big(M; \boldsymbol{Q}_l \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) + \underbrace{\mathsf{I}\big(M; \boldsymbol{A}_l \,|\, \boldsymbol{Q}_l, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big)}_{=0} \\
& = \mathsf{I}\big(M; \boldsymbol{Q}_l\big) \leq \rho_l.
\end{aligned}
$$

Hence, (39) follows from Lemmas 4 and 5. In addition, (39) implies that

$$
\begin{aligned}
\mathsf{H}\big(\boldsymbol{A}_l^{(m)} \,|\, \boldsymbol{Q}_l^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) & \\
\geq \mathsf{H}\big(\boldsymbol{A}_l^{(m')} \,|\, \boldsymbol{Q}_l^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) & \\
- \big[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\big], &
\end{aligned}
\tag{40}
$$

for all $l \in [n]$.

Now, due to (1) we know that

$$
\begin{aligned}
& \mathsf{H}\big(\boldsymbol{A}_{[n]}^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}\big) \\
& = \mathsf{H}\big(\boldsymbol{X}^{(m)}, \boldsymbol{A}_{[n]}^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}\big) \\
& \quad - \underbrace{\mathsf{H}\big(\boldsymbol{X}^{(m)} \,|\, \boldsymbol{A}_{[n]}^{(m)}, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}\big)}_{=0} \\
& = \mathsf{H}\big(\boldsymbol{X}^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}\big) + \mathsf{H}\big(\boldsymbol{A}_{[n]}^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& = \beta \log_2 |\mathcal{X}| + \mathsf{H}\big(\boldsymbol{A}_{[n]}^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& \quad - \underbrace{\mathsf{H}\big(\boldsymbol{A}_{[n]}^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, \boldsymbol{X}^{\mathcal{M}^{\mathrm{c}} \backslash \{m\}}\big)}_{=0} \\
& = \beta \log_2 |\mathcal{X}| + \mathsf{I}\big(\boldsymbol{A}_{[n]}^{(m)}; \boldsymbol{X}^{\mathcal{M}^{\mathrm{c}} \backslash \{m\}} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& = \beta \log_2 |\mathcal{X}| \\
& \quad + \mathsf{I}\big(\boldsymbol{A}_{[n]}^{(m)}, \boldsymbol{Q}_{[n]}^{(m)}; \boldsymbol{X}^{\mathcal{M}^{\mathrm{c}} \backslash \{m\}} \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \tag{41} \\
& \geq \beta \log_2 |\mathcal{X}| \\
& \quad + \mathsf{I}\big(\boldsymbol{A}_l^{(m)}, \boldsymbol{Q}_{[n]}^{(m)}; \boldsymbol{X}^{\mathcal{M}^{\mathrm{c}} \backslash \{m\}} \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \tag{42} \\
& = \beta \log_2 |\mathcal{X}| + \mathsf{H}\big(\boldsymbol{A}_l^{(m)}, \boldsymbol{Q}_{[n]}^{(m)} \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& \quad - \mathsf{H}\big(\boldsymbol{A}_l^{(m)}, \boldsymbol{Q}_{[n]}^{(m)} \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, \boldsymbol{X}^{\mathcal{M}^{\mathrm{c}} \backslash \{m\}}\big) \\
& = \beta \log_2 |\mathcal{X}| + \underbrace{\mathsf{H}\big(\boldsymbol{Q}_{[n]}^{(m)} \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big)}_{= \mathsf{H}\big(\boldsymbol{Q}_{[n]}^{(m)}\big)} \\
& \quad + \mathsf{H}\big(\boldsymbol{A}_l^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& \quad - \underbrace{\mathsf{H}\big(\boldsymbol{Q}_{[n]}^{(m)} \,|\, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, \boldsymbol{X}^{\mathcal{M}^{\mathrm{c}} \backslash \{m\}}\big)}_{= \mathsf{H}\big(\boldsymbol{Q}_{[n]}^{(m)}\big)} \\
& \quad - \underbrace{\mathsf{H}\big(\boldsymbol{A}_l^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}, \boldsymbol{X}^{\mathcal{M}^{\mathrm{c}} \backslash \{m\}}\big)}_{=0} \\
& = \beta \log_2 |\mathcal{X}| + \mathsf{H}\big(\boldsymbol{A}_l^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& = \beta \log_2 |\mathcal{X}| + \mathsf{H}\big(\boldsymbol{A}_l^{(m)} \,|\, \boldsymbol{Q}_l^{(m)}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \tag{43} \\
& \geq \beta \log_2 |\mathcal{X}| + \mathsf{H}\big(\boldsymbol{A}_l^{(m')} \,|\, \boldsymbol{Q}_l^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& \quad - \big[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\big] \tag{44} \\
& = \beta \log_2 |\mathcal{X}| + \mathsf{H}\big(\boldsymbol{A}_l^{(m')} \,|\, \boldsymbol{Q}_{[n]}^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& \quad - \big[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\big], \tag{45}
\end{aligned}
$$

for any $l \in [n]$, where (41) follows from (2), (42) holds by the chain rule for MI, and the final inequality (44) is due to (40).

Thus, summing (45) over all possible $l \in [n]$ we have

$$
\begin{aligned}
& n \, \mathsf{H}\big(\boldsymbol{A}_{[n]}^{(m)} \,|\, \boldsymbol{Q}_{[n]}^{(m)}, \boldsymbol{X}^{\mathcal{M}}\big) \\
& \geq n \beta \log_2 |\mathcal{X}| + \sum_{l=1}^{n} \mathsf{H}\big(\boldsymbol{A}_l^{(m')} \,|\, \boldsymbol{Q}_{[n]}^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& \quad - \sum_{l=1}^{n} \big[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\big] \\
& \geq n \beta \log_2 |\mathcal{X}| + \mathsf{H}\big(\boldsymbol{A}_{[n]}^{(m')} \,|\, \boldsymbol{Q}_{[n]}^{(m')}, \boldsymbol{X}^{\mathcal{M}}, \boldsymbol{X}^{(m)}\big) \\
& \quad - \sum_{l=1}^{n} \big[\epsilon^{\mathsf{MI}}(\mathcal{Q}_l, \mathcal{A}_l) + \epsilon^{\mathsf{MI}}(\mathcal{Q}_l)\big].
\end{aligned}
$$

The result of (23) then follows by dividing both sides by n.

On the other hand, following a similar derivation as (43), we can obtain

$$
\begin{aligned}
& \mathsf{H}\big(\boldsymbol{A}_{[n]}^{(\mathrm{M})} \,|\, \boldsymbol{Q}_{[n]}^{(\mathrm{M})}, \boldsymbol{X}^{[\mathrm{M}-1]}\big) \\
& \geq \beta \log_2 |\mathcal{X}| + \underbrace{\mathsf{H}\big(\boldsymbol{A}_l^{(\mathrm{M})} \,|\, \boldsymbol{Q}_l^{(\mathrm{M})}, \boldsymbol{X}^{(\mathrm{M})}, \boldsymbol{X}^{[\mathrm{M}-1]}\big)}_{=0}.
\end{aligned}
$$

This completes the proof of (24).

APPENDIX H
PROOF OF LEMMA 7

We start the proof by defining a set $\mathcal{B}_{x,x'} \triangleq \{x, x'\}$ with two arbitrary elements $x \neq x'$, $x, x' \in \mathcal{X}$, and a subset $\mathcal{Z}_{x,x'} \subseteq \mathcal{Y}$ as $\{y \in \mathcal{Y} \colon P_{Y|X}(y|x) \geq P_{Y|X}(y|x')\}$. Next, we introduce a new RV $Z_{x,x'}$ as

$$
Z_{x,x'}(y) = \begin{cases} 1 & \text{if } y \in \mathcal{Z}_{x,x'}, \\ 0 & \text{otherwise.} \end{cases}
$$

By (5) we have

$$
\begin{aligned}
& \sum_{z \in \{0,1\}} \max_{b \in \mathcal{B}_{x,x'}} P_{Z_{x,x'}|X}(z|b) \\
& = \sum_{y \in \mathcal{Z}_{x,x'}} P_{Y|X}(y|x) + \sum_{y \in \mathcal{Z}_{x,x'}^{\mathrm{c}}} P_{Y|X}(y|x') \\
& \overset{(a)}{=} \sum_{y \in \mathcal{Y}} \max_{b \in \mathcal{B}_{x,x'}} P_{Y|X}(y|b) \\
& \overset{(b)}{\leq} \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{Y|X}(y|x) \leq 2^{\rho},
\end{aligned}
$$

where (a) follows by the definition of the subset $\mathcal{Z}_{x,x'}$, and (b) holds simply because maximizing over a subset leads to a smaller value. Moreover, by using the relation between TV distance and variational distance [75, Lem. 3.12], it follows

that

$$\left\| P_{Z_{x,x'}|X=x} - P_{Z_{x,x'}|X=x'} \right\|_{\mathsf{TV}}$$

$$= \frac{1}{2}\left(\left| P_{Z_{x,x'}|X}(1|x) - P_{Z_{x,x'}|X}(1|x') \right| \right.$$

$$\left. + \left| P_{Z_{x,x'}|X}(0|x) - P_{Z_{x,x'}|X}(0|x') \right| \right)$$

$$= \frac{1}{2}\left(\sum_{y \in \mathcal{Z}_{x,x'}} \left[P_{Y|X=x}(y) - P_{Y|X=x'}(y) \right] \right.$$

$$\left. + \sum_{y \in \mathcal{Z}^{\mathsf{c}}_{x,x'}} \left[P_{Y|X=x'}(y) - P_{Y|X=x}(y) \right] \right)$$

$$= \frac{1}{2} \sum_{y \in \mathcal{Y}} \left| P_{Y|X=x}(y) - P_{Y|X=x'}(y) \right|$$

$$= \left\| P_{Y|X=x} - P_{Y|X=x'} \right\|_{\mathsf{TV}}.$$

Therefore, since x, x' are chosen arbitrarily, Lemma 7 holds for any alphabets $\mathcal{X}$ and $\mathcal{Y}$ if we can show that the assertion is true when X and Y are binary.

To complete the proof, we show that Lemma 7 holds for any binary RVs X and Y. The proof is quite straightforward: Since $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, we can define $P_{Y|X}$ by $a \triangleq P_{Y|X}(1|0)$ and $b \triangleq P_{Y|X}(1|1)$ with $0 \leq a, b \leq 1$. Thus, by definition $\|P_{Y|X=0} - P_{Y|X=1}\|_{\mathsf{TV}} = |a - b|$ and $\mathsf{MaxL}(X; Y) = \log_2\left(\max\{1-a, 1-b\} + \max\{a, b\}\right) = \log_2\left(1 + |a - b|\right) \leq \rho$. As $\log_2(\cdot)$ is a strictly increasing function, $|a - b| \leq 2^\rho - 1$.

ACKNOWLEDGMENT

REFERENCES

[1] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proc. 36th Annu. IEEE Symp. Found. Comp. Sci. (FOCS)*, Milwaukee, WI, USA, Oct. 1995, pp. 41–50.

[2] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," *J. ACM*, vol. 45, no. 6, pp. 965–982, Nov. 1998.

[3] A. Beimel, Y. Ishai, E. Kushilevitz, and J.-F. Raymond, "Breaking the $O(n^{1/(2k-1)})$ barrier for information-theoretic private information retrieval," in *Proc. 43rd Annu. IEEE Symp. Found. Comp. Sci. (FOCS)*, Vancouver, BC, Canada, Nov. 2002, pp. 261–270.

[4] Z. Dvir and S. Gopi, "2-server PIR with subpolynomial communication," *J. ACM*, vol. 63, no. 4, pp. 39:1–39:15, Nov. 2016.

[5] S. Yekhanin, "Towards 3-query locally decodable codes of subexponential length," *J. ACM*, vol. 55, no. 1, pp. 1:1–1:16, Feb. 2008.

[6] H. Corrigan-Gibbs and D. Kogan, "Private information retrieval with sublinear online time," in *Proc. 39th Annu. Int. Conf. Theory Appl. Crypto. Techn. (EUROCRYPT)*, Zagreb, Croatia, May 2020, pp. 44–75.

[7] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, Jul. 2017.

[8] N. B. Shah, K. V. Rashmi, and K. Ramchandran, "One extra bit of download ensures perfectly private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 856–860.

[9] T. H. Chan, S.-W. Ho, and H. Yamamoto, "Private information retrieval for coded storage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 2842–2846.

[10] A. Fazeli, A. Vardy, and E. Yaakobi, "Codes for distributed PIR with low storage overhead," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 2852–2856.

[11] R. Tajeddine, O. W. Gnilke, and S. El Rouayheb, "Private information retrieval from MDS coded data in distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7081–7093, Nov. 2018.

[12] K. Banawan and S. Ulukus, "The capacity of private information retrieval from coded databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1945–1956, Mar. 2018.

[13] H. Sun and C. Tian, "Breaking the MDS-PIR capacity barrier via joint storage coding," *Information*, vol. 10, no. 9, Aug. 2019.

[14] J. Zhu, Q. Yan, C. Qi, and X. Tang, "A new capacity-achieving private information retrieval scheme with (almost) optimal file length for coded servers," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1248–1260, 2020.

[15] R. Zhou, C. Tian, H. Sun, and T. Liu, "Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4904–4916, Aug. 2020.

[16] J. Li, D. Karpuk, and C. Hollanti, "Towards practical private information retrieval from MDS array codes," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3415–3425, Jun. 2020.

[17] S. Kumar, H.-Y. Lin, E. Rosnes, and A. Graell i Amat, "Achieving maximum distance separable private information retrieval capacity with linear codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4243–4273, Jul. 2019.

[18] H.-Y. Lin, S. Kumar, E. Rosnes, and A. Graell i Amat, "Asymmetry helps: Improved private information retrieval protocols for distributed storage," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Guangzhou, China, Nov. 2018, pp. 1–5.

[19] R. Freij-Hollanti, O. W. Gnilke, C. Hollanti, A.-L. Horlemann-Trautmann, D. Karpuk, and I. Kubjas, "t-private information retrieval schemes using transitive codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2107–2118, Apr. 2019.

[20] J. Lavauzelle, R. Tajeddine, R. Freij-Hollanti, and C. Hollanti, "Private information retrieval schemes with product-matrix MBR codes," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 441–450, 2021.

[21] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2361–2370, Apr. 2018.

[22] R. Freij-Hollanti, O. W. Gnilke, C. Hollanti, and D. A. Karpuk, "Private information retrieval from coded databases with colluding servers," *SIAM J. Appl. Algebra Geometry*, vol. 1, no. 1, pp. 647–664, Nov. 2017.

[23] R. G. L. D'Oliveira and S. El Rouayheb, "One-shot PIR: Refinement and lifting," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2443–2455, Apr. 2020.

[24] L. Holzbaur, R. Freij-Hollanti, J. Li, and C. Hollanti, "Towards the capacity of private information retrieval from coded and colluding servers," *IEEE Trans. Inf. Theory*, early access, Oct. 14, 2021, doi: 10.1109/TIT.2021.3120316.

[25] K. Banawan and S. Ulukus, "The capacity of private information retrieval from Byzantine and colluding databases," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1206–1219, Feb. 2019.

[26] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, and C. Hollanti, "Private information retrieval from coded storage systems with colluding, Byzantine, and unresponsive servers," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3898–3906, Jun. 2019.

[27] H. Sun and S. A. Jafar, "Multiround private information retrieval: Capacity and storage overhead," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5743–5754, Aug. 2018.

[28] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *IEEE Trans. Inf. Theory*, vol. 64, no. 10, pp. 6842–6862, Oct. 2018.

[29] H. Sun and S. A. Jafar, "Optimal download cost of private information retrieval for arbitrary message length," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 12, pp. 2920–2932, Dec. 2017.

[30] C. Tian, H. Sun, and J. Chen, "Capacity-achieving private information retrieval codes with optimal message size and upload cost," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7613–7627, Nov. 2019.

[31] Y. Zhang, E. Yaakobi, T. Etzion, and M. Schwartz, "On the access complexity of PIR schemes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 2134–2138.

[32] C. Tian, "On the storage cost of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7539–7549, Dec. 2020.

[33] R. Tandon, "The capacity of cache aided private information retrieval," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2017, pp. 1078–1082.

[34] Y.-P. Wei, K. Banawan, and S. Ulukus, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3215–3232, May 2019.

[35] S. Li and M. Gastpar, "Single-server multi-user private information retrieval with side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 1954–1958.

[36] Y.-P. Wei, K. Banawan, and S. Ulukus, "The capacity of private information retrieval with partially known private side information," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8222–8231, Dec. 2019.

[37] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson, "Private information retrieval with side information," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2032–2043, Apr. 2020.

[38] Z. Chen, Z. Wang, and S. A. Jafar, "The capacity of T-private information retrieval with private side information," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4761–4773, Aug. 2020.

[39] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "The role of coded side information in single-server private information retrieval," *IEEE Trans. Inf. Theory*, vol. 67, no. 1, pp. 25–44, Jan. 2021.

[40] N. Raviv, I. Tamo, and E. Yaakobi, "Private information retrieval in graph-based replication systems," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3590–3602, Jun. 2020.

[41] Z. Jia and S. A. Jafar, "On the asymptotic capacity of X-secure T-private information retrieval with graph-based replicated storage," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6280–6296, Oct. 2020.

[42] H. Yang, W. Shin, and J. Lee, "Private information retrieval for secure distributed storage systems," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 12, pp. 2953–2964, Dec. 2018.

[43] Z. Jia, H. Sun, and S. A. Jafar, "Cross subspace alignment and the asymptotic capacity of X-secure T-private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5783–5798, Sep. 2019.

[44] Z. Jia and S. A. Jafar, "X-secure T-private information retrieval from MDS coded storage with Byzantine and unresponsive servers," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7427–7438, Dec. 2020.

[45] Y. Zhang, T. Etzion, and E. Yaakobi, "Bounds on the length of functional PIR and batch codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4917–4934, Aug. 2020.

[46] Y. Zhang, E. Yaakobi, and T. Etzion, "Private proximity retrieval codes," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7458–7476, Nov. 2021.

[47] D. Asonov and J.-C. Freytag, "Repudiative information retrieval," in *Proc. ACM Workshop Privacy Electron. Soc. (WPES)*, Washington, DC, USA, Nov. 2002, pp. 32–40.

[48] R. R. Toledo, G. Danezis, and I. Goldberg, "Lower-cost ϵ-private information retrieval," in *Proc. Privacy Enhancing Technol. Symp. (PETS)*, Darmstadt, Germany, Jul. 2016, pp. 184–201.

[49] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptography Conf. (TCC)*, New York, NY, USA, Mar. 2006, pp. 265–284.

[50] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Coll. Automata, Lang. Program. (ICALP), Part II*, Venice, Italy, Jul. 2006, pp. 1–12.

[51] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 57:1–57:38, Jul. 2018.

[52] Z. Lin, M. Hewett, and R. B. Altman, "Using binning to maintain confidentiality of medical data," in *Proc. Amer. Med. Inform. Assoc. Symp. (AMIA)*, San Antonio, TX, USA, Nov. 2002, pp. 454–458.

[53] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 838–852, Jun. 2013.

[54] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *Proc. 23rd ACM SIGSAC Conf. Comput. Commun. Security (CCS)*, Vienna, Austria, Oct. 2016, pp. 43–54.

[55] G. Smith, "On the foundations of quantitative information flow," in *Proc. 12th Int. Conf. Found. Softw. Sci. Comput. Struct. (FoSSaCS)*, York, U.K., Mar. 2009, pp. 288–302.

[56] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1625–1657, Mar. 2020.

[57] B. Köpf and D. Basin, "An information-theoretic model for adaptive side-channel attacks," in *Proc. 14th ACM SIGSAC Conf. Comput. Commun. Security (CCS)*, Alexandria, VA, USA, Oct. 2007, pp. 286–296.

[58] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" in *Proc. 49th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Philadelphia, PA, USA, Oct. 2008, pp. 531–540.

[59] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. 54th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Berkeley, CA, USA, Oct. 2013, pp. 429–438.

[60] I. Samy, R. Tandon, and L. Lazos, "On the capacity of leaky private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1262–1266.

[61] I. Samy, M. Attia, R. Tandon, and L. Lazos, "Asymmetric leaky private information retrieval," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 5352–5369, Aug. 2021.

[62] I. Samy, M. A. Attia, R. Tandon, and L. Lazos, "Latent-variable private information retrieval," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angles, CA, USA, Jun. 2020, pp. 1071–1076.

[63] H. Sun and S. A. Jafar, "The capacity of symmetric private information retrieval," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 322–329, Jan. 2019.

[64] Q. Wang and M. Skoglund, "On PIR and symmetric PIR from colluding databases with adversaries and eavesdroppers," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3183–3197, May 2019.

[65] Q. Wang and M. Skoglund, "Symmetric private information retrieval from MDS coded distributed storage with non-colluding and colluding servers," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5160–5175, Aug. 2019.

[66] R. Zhou, T. Guo, and C. Tian, "Weakly private information retrieval under the maximal leakage metric," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angles, CA, USA, Jun. 2020, pp. 1089–1094.

[67] H.-Y. Lin, S. Kumar, E. Rosnes, A. Graell i Amat, and E. Yaakobi, "The capacity of single-server weakly-private information retrieval," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 415–427, Mar. 2021.

[68] T. Guo, R. Zhou, and C. Tian, "On the information leakage in private information retrieval systems," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2999–3012, 2020.

[69] S. Vithana, K. Banawan, and S. Ulukus, "Semantic private information retrieval: Effects of heterogeneous message sizes and popularities," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Taipei, Taiwan, Dec. 2020, pp. 1–6.

[70] G. Barthe and B. Köpf, "Information-theoretic bounds for differentially private mechanisms," in *Proc. 24th IEEE Comput. Security Found. Symp. (CSF)*, Cernay-la-Ville, France, Jun. 2011, pp. 191–204.

[71] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Berkeley, CA, USA: Univ. California Press, 1961, pp. 547–561.

[72] C. Tian, H. Sun, and J. Chen, "A Shannon-theoretic approach to the storage-retrieval tradeoff in PIR systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 1904–1908.

[73] Z. Zhang, "Estimating mutual information via Kolmogorov distance," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3280–3282, Sep. 2007.

[74] S.-W. Ho and R. W. Yeung, "The interplay between entropy and variational distance," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 5906–5929, Dec. 2010.

[75] S. M. Moser, "Advanced topics in information theory (lecture notes)," 4th ed., Signal Inf. Process. Lab., ETH Zürich, Switzerland, Dept. Elect. Comput. Eng., Nat. Chiao Tung Univ., Hsinchu, Taiwan, 2019. [Online]. Available: http://moser-isi.ethz.ch/scripts.html

[76] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2909–2913, Jan. 2016.

[77] A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd, "A rewriting system for convex optimization problems," *J. Control Decision*, vol. 5, no. 1, pp. 42–60, Jan. 2018.

[78] Z. Chen, Z. Wang, and S. A. Jafar, "The asymptotic capacity of private search," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 4709–4721, Aug. 2020.

Hsuan-Yin Lin (Senior Member, IEEE) received the B.S. degree from the National Tsing-Hua University (NTHU), Taiwan, in 2007, and the M.S. and Ph.D. degrees from the National Chiao Tung University (NCTU), Taiwan, in 2008 and 2013, respectively. He was a Visiting Scholar at Universitat Pompeu Fabra, Barcelona, Spain, and TU Darmstadt, Germany, and held a post-doctoral position at Simula UiB, Norway, where he is currently a Research Scientist. His current interests include privacy-preserving technologies, information-theoretic cryptography, coding in distributed storage systems, finite blocklength information theory, scheduling in millimeter-wave cellular networks, and distributed detection and estimation. In 2014, he was awarded the Honor Membership of the Phi Tau Phi Scholastic Honor Society of the Republic of China (Taiwan) and the New Partnership Program for the Connection to the Top Labs in the World (subsidized by the Ministry of Science and Technology, Taiwan). He was the Publicity Chair of the 9th International Workshop on Signal Design and its Applications in Communications, Dongguan, China. He is currently serving as a Guest Editor for *Entropy*: Special Issue "Information-theoretic approach to privacy and security," and as a TPC member of the 2022 International Zurich Seminar on Information and Communication (IZS).

Siddhartha Kumar received the M.Sc. degree in communications engineering from the Chalmers University of Technology, Gothenburg, Sweden, in 2015, and the Ph.D. degree in informatics from the University of Bergen, Bergen, Norway, in 2018. He is currently a Post-Doctoral Fellow at Simula UiB. He has been a TPC Member of the IEEE Global Communications Conference (GLOBECOM) in 2020 and 2021, and the IEEE Wireless Communications and Networking Conference (WCNC) in 2021. His research interests are geared toward information-theoretic privacy and security in distributed storage and computing. In 2013, his bachelor's thesis was funded by the Karnataka Government, India, under the KSCST Student Project Program.

Eirik Rosnes (Senior Member, IEEE) was born in Stavanger, Norway, in 1975. He received the Cand. Scient. degree in physics and the Dr. Scient. degree in informatics from the University of Bergen, Bergen, Norway, in 1999 and 2003, respectively. From 2003 to 2011, he was with the Department of Informatics, University of Bergen, first as a Post-Doctoral Researcher and then as a Senior Researcher. From September 2001 to March 2002 and from March 2005 to September 2005, he was a Visiting Scholar with the Center for Magnetic Recording Research, University of California at San Diego, La Jolla, CA, USA. From October 2011 to August 2013, he was a Senior Engineer with Ceragon Networks, and since August 2013, he has been with both Simula Research Laboratory/Simula UiB as a Senior Researcher (since 2020 as a Chief Research Scientist) and Section Leader and the University of Bergen (until August 2017) as an Adjunct Associate Professor. His research interests are in the areas of communication theory and information theory and include classical error-control coding, codes on graphs, codes for distributed storage systems, and coding for privacy and security. He was the Technical Program Co-Chair of the 2009 International Workshop on Coding and Cryptography, Ullensvang, Norway, and served as an Area Editor for the *AEÜ—International Journal of Electronics and Communications* from October 2014 to January 2020. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS.

Alexandre Graell i Amat (Senior Member, IEEE) received the M.Sc. degree in telecommunications engineering from the Universitat Politècnica de Catalunya, Barcelona, Catalonia, Spain, in 2001, and the M.Sc. and Ph.D. degrees in electrical engineering from the Politecnico di Torino, Turin, Italy, in 2000 and 2004, respectively.

From 2001 to 2002, he was a Visiting Scholar with the University of California at San Diego, La Jolla, CA, USA. From 2002 to 2003, he held a visiting appointment at Universitat Pompeu Fabra, Barcelona, and the Telecommunications Technological Center of Catalonia, Barcelona. From 2001 to 2004, he held a part-time appointment at STMicroelectronics Data Storage Division, Milan, Italy, as a Consultant on coding for magnetic recording channels. From 2004 to 2005, he was a Visiting Professor with Universitat Pompeu Fabra. From 2006 to 2010, he was with the Department of Electronics, IMT Atlantique (formerly ENST Bretagne), Brest, France. Since 2019, he has also been an Adjunct Research Scientist at Simula UiB, Bergen, Norway. He is currently a Professor with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. His research interests are in the field of coding theory with application to distributed computing, privacy and security, random access, and optical communications. He received the Marie Skłodowska-Curie Fellowship from the European Commission and the Juan de la Cierva Fellowship from the Spanish Ministry of Education and Science. He received the IEEE Communications Society 2010 Europe, Middle East, and Africa Region Outstanding Young Researcher Award. He was the General Co-Chair of the 7th International Symposium on Turbo Codes and Iterative Information Processing, Sweden, in 2012, and the TPC Co-Chair of the 11th International Symposium on Topics in Coding, Canada, in 2021. He was an Associate Editor of the IEEE COMMUNICATIONS LETTERS from 2011 to 2013. He was an Associate Editor and the Editor-at-Large of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2011 to 2016 and 2017 to 2020, respectively. He is currently an Area Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS.

Eitan Yaakobi (Senior Member, IEEE) received the B.A. degree in computer science and mathematics and the M.Sc. degree in computer science from the Technion—Israel Institute of Technology, Haifa, Israel, in 2005 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of California at San Diego in 2011. From 2011 to 2013, he was a Post-Doctoral Researcher with the Department of Electrical Engineering, California Institute of Technology, and with the Center for Memory and Recording Research, University of California at San Diego. He is currently an Associate Professor with the Department of Computer Science, Technion—Israel Institute of Technology. His research interests include information and coding theory with applications to non-volatile memories, associative memories, DNA storage, data storage and retrieval, and private information retrieval. He received the Marconi Society Young Scholar in 2009 and the Intel Ph.D. Fellowship in 2010–2011.